Identification and Estimation of Finite Mixtures of Multinomial Logit Models

Dingyi Li

Cornell University

September 15, 2025

Link for the latest version

Abstract

Finite mixtures of multinomial logit models can be used to capture consumer choice heterogeneity across multiple markets when only aggregate consumer choices per market are available. A motivating example is a nested logit where the composition of each mixture component (each nest of alternatives) is unknown a priori. We show that in order to identify these models, it suffices to require that each mixture component includes at least two component-exclusive alternatives. We refer to our assumption as the *pure-alternatives* condition, and we argue it is a natural extension of the *anchor-word* assumption used commonly in nonnegative matrix factorization problems in machine learning. Our identification result enables a consistent two-step estimator as the number of consumers, markets, and alternatives grow large. Applying this framework to the U.S. vehicle market, we find that consumer heterogeneity does not yield substitution patterns between electric and internal combustion engine vehicles, suggesting consumer segments are distinctly aligned with specific vehicle types without crossover substitution.

Keywords: discrete choice, finite mixtures, machine learning.

1 Introduction

In this paper, we study Finite Mixtures of Multinomial Logit Models (FMML). These are categorical finite mixture models, where each mixture component arises from a multinomial discrete choice model with observed and unobserved endogenous characteristics and an idiosyncratic error term that follows a standard Type-1 extreme value distribution. The goal of the FMML model is to identify the mixing distributions (the proportions of subpopulations that share mixture components) and to estimate the corresponding preference parameters in each of the finitely many mixture components. Unlike mixed logit methods, which assume a known distribution of subpopulations maximizing utility across all alternatives, FMML uncovers latent subpopulations with unknown memberships who may optimize utility over an unknown subset of alternatives. This provides a powerful framework for analyzing hidden heterogeneity in choice behavior.

Mixture models are widely used in economics and other fields (Compiani and Kitamura, 2016; McLachlan et al., 2019), offering solutions when a single distribution cannot capture the complexity of the observed data. Finite mixtures, particularly, provide a simple model for analyzing complex data structures. However, identifying and estimating these models in our setting presents two important challenges. First, only aggregate data, such as market shares, are observed rather than individual choices. Second, membership in latent subpopulations is unobserved, complicating the assignment of choices to specific mixtures. In this paper, we provide identification criteria and estimation strategies for FMML models to address these issues.

In this paper, we make two contributions to the literature. First, we show that the existence of *pure alternatives*—alternatives that are specific to a mixture component—can be used to identify the parameters of the model, even when no individual-level data are available. Second, we introduce a novel two-step procedure for estimating Finite Mixtures of Logit Models (FMML) using aggregate data.

The two-step procedure begins by assuming the existence of instrumental variables that shift the observed endogenous variables. Then, the variation generated by the instrumental variables—along with the pure alternatives assumption and the multinomial logit structure—ensures that the log-share difference between pairs of pure alternatives *within the same mixture component* satisfy certain linearity assumption (as a function of the differences in the observed endogenous variables). Consequently, a nonparametric functional form test can be used to estimate the identity of the pure alternatives. The nonparametric test we use in the paper is based on the work of Fan and Li (1996) and leverages the Independence of Irrelevant Alternatives (IIA) property in each mixture component. Novel use of concentration inequalities helps derive a tuning parameter that guarantees that pure alternatives can be selected with high probability. The nonparametric test is applied to the residuals of *pairwise* regressions in which the outcome variable is the log-share difference between a pair of alternatives, and the regressor of interest is the endogenous variable. These regressions are estimated by two-stage least squares (2SLS). In the second stage, the suggested estimation procedure recovers mixture weights using the identified pure alternatives.

The identification assumption proposed in this paper is an extension of the *anchor-word* assumption (or more generally the *separability condition*) used commonly in nonnegative matrix factorization problems in machine learning (Donoho and Stodden, 2003; Arora et al., 2012a,b). In the context of topic models for text data, anchor words are unique terms associated exclusively with specific topics (Bing et al., 2020b,c; Ke and Wang, 2022).[1] In our model, pure alternatives are alternatives that are specific to each mixture component.

A notable feature of FMML is that—beyond the pure alternatives and the multinomial logit structure—its estimation does not require any additional substantial information about the mixture components. In the context of nested logits (which are a particular case of the FMML), this means that it is not necessary to pre-specify the nesting of alternatives, a major departure from traditional models (Nevo, 1998; Train, 2009; Miller et al., 2021; Fosgerau et al., 2024). We also think that the FMML model could be useful in the context of the literature on consideration sets (Manski, 1977; Swait and Ben-Akiva, 1987; Abaluck et al., 2020; Barseghyan et al., 2021a,b; Barseghyan and Molinari, 2023; Abaluck and Adams-Prassl, 2021; Agarwal and Somaini, 2022). Broadly speaking, this literature studies models where decision-makers consider only a subset of the available alternatives when making choices. While traditional logit models assume that all available options are considered, decision-makers often evaluate only a smaller subset (this subset is called the consideration set). The identification arguments used in this paper are different to the ones currently used in this literature. Moreover, theoretically speaking, our framework can allow for the consideration sets to depend on observed (exogenous and endogenous) covariates and also unobserved characteristics.

We demonstrate the efficacy of this estimation approach using a Monte Carlo simulation and then apply the method to an empirical application: estimating demand in the US electric vehicle market using yearly data from 2011 to 2020. The algorithm classifies two groups: Electric Vehicles (EVs) and Internal Combustion Engine (ICE) vehicles, representing two pure alternative pairs. Results reveal no significant substitution pattern between EVs and ICEs, with substitution declining further when incorporating second-choice data, as observed by Xing et al. (2021). These findings support the consideration set literature over traditional nested logit models, which typically assume positive substitution patterns among nests.

[1] The separability assumption is also used in other fields beyond the analysis of text data—for example, the "pure-node" assumption in community detection (Airoldi et al., 2008; Mao et al., 2017) and the "pure pixel" assumption in hyperspectral imaging (Ma et al., 2013). This shows the broad applicability of separability concepts across diverse contexts.

This paper also connects to machine learning approaches that incorporate sparsity-like assumptions in economic models. Unlike post-LASSO methods (Belloni et al., 2012), which use ℓ_1 penalties, the clustering method proposed here recovers sparse signals while enhancing computational efficiency and interpretability. Using pairwise comparisons, our approach improves the computational complexity of Bonhomme and Manresa (2015), capable of dealing with nested logit models. In contrast, our method accommodates overlapping structures and an unknown number of types. The identified sparse signals serve as anchors, enabling the recovery of high-dimensional parameters within logit models featuring finite mixtures. Furthermore, the stepwise procedure (Kozbur, 2017; Bing et al., 2020a) contrasts with the EM algorithm, yielding interpretable results due to its inherent sparsity. This is particularly evident when pure alternatives are identified, offering a straightforward and transparent framework for understanding the mixture components and their underlying data structures.

The paper is organized as follows: Section 2 discusses the settings of FMML, Section 3 presents the model assumptions and identification strategy, Section 4 details the estimation methods, Section 5 offers empirical evidence, and Section 6 explores model extensions. Section 7 concludes, with detailed proofs provided in the appendix.

2 Setting

2.1 General notation

The following notation will be used in the paper. The set $\{1, \ldots, n\}$ is denoted by $[n]$. For a generic set S, we define $|S|$ to be its cardinality. For a generic vector $v \in \mathbb{R}^d$, we denote $\|v\|_q$ to be the vector's l_q norm for $q = 0, 1, 2, \ldots, \infty$, and $\text{diag}(v)$ to be a $d \times d$ diagonal matrix with diagonal elements equal to v.

2.2 Model

A decision maker n in market $t \in [T]$ selects a single good from a set of alternatives $j \in [J]$. In contrast to the framework in McFadden (1972) and Berry (1994), the decision maker is characterized by a random type $\theta \in [K]$. A decision maker with type θ only chooses alternatives belonging to $J_\theta \subseteq [J]$.[2] As is common in applications, we assume decision makers observe market-level characteristics

$$\mathcal{X}_t := (x_t, p_t, \xi_t^u)$$

[2]The decision maker may either consciously select this nest or automatically narrow her choices based on limited consideration. In the second stage, she makes a final choice from alternatives within this selected subset.

where $x_t = (x_{1t}, \dots, x_{Jt})$ represents exogenous observed characteristics with $x_{jt} \in \mathbb{R}^{d-1}$, $p_t = (p_{1t}, \dots, p_{Jt})$ denotes endogenous observed characteristics with $p_{jt} \in \mathbb{R}$, and $\xi_t^u = (\xi_{1t}^u, \dots, \xi_{Kt}^u)$ captures unobserved type-market characteristics, with each $\xi_{kt}^u = (\xi_{k1t}^u, \dots, \xi_{kJt}^u)$ where $\xi_{kjt}^u \in \mathbb{R}$ for $k \in [K]$. We refer to p_{jt} as the price correlated with the structural errors ξ_t^u.[3] The superscript u indicates that these unobserved characteristics have not been normalized to have mean zero through fixed effects. When $d = 1$, we allow for no observed exogenous covariates.

The conditional indirect utility of a decision maker n from choosing alternative j from nest J_θ in market t is represented as $v_{n\theta jt}$ where we adopt the multinomial logit model, given by

$$v_{n\theta jt} := \underbrace{x_{jt}^\top \beta_\theta - \alpha_\theta p_{jt} + u_{\theta j} + \lambda_\theta \xi_{jt}}_{\delta_{\theta jt}} + \epsilon_{n\theta jt}$$

with heterogeneous coefficients $\gamma_\theta = (\beta_\theta, \alpha_\theta) \in B^d$ in the different mixtures.[4,5] $\lambda_\theta \in \Lambda$ is a user-specific/estimated variable explaining the relative utility of the unobserved quality, ξ_{jt}. The idiosyncratic preference shock to decision maker n, denoted as $\epsilon_{n\theta jt}$, is assumed to follow an independent Type I extreme value distribution.[6] The mean utility of product j for consumers of type θ is $\delta_{\theta jt}$ and define $\delta_{\theta t} := (\delta_{\theta 1t}, \cdots, \delta_{\theta Jt})$.

The choice probability of product j in market t, denoted by s_{jt}, represents a single consumer's choice behavior, averaged over consumer types, idiosyncratic shocks, and other factors not captured in $\mathcal{X}_t$. Derived from utility maximization, s_{jt} is given by

$$\begin{aligned} s_{jt} :=& \mathbb{E}\left(s_{njt} \mid \mathcal{X}_t\right) = \mathbb{E}\left[\mathbb{E}\left(s_{njt} \mid \theta, \mathcal{X}_t\right) \mid \mathcal{X}_t\right] \\ =& \sum_{k \in [K]} \mathbb{P}(\theta = k \mid \mathcal{X}_t) \mathbb{P}\left(\operatorname{argmax}_{i \in J_\theta} v_{n\theta it} = j \mid \theta = k, \delta_{\theta t}\right), \end{aligned} \tag{1}$$

where $s_{njt}(\theta, \mathcal{X}_t) := \mathbb{1}\{\operatorname{argmax}_{i \in J_\theta} v_{n\theta it} = j\}$ abbreviated here as s_{njt} and $\mathbb{P}(\theta = k \mid \mathcal{X}_t)$ represents the probability distribution of consumer types in market t, assumed homogeneous across consumers. The choice probabilities with simplified notation are

$$s_{jt} = \sum_{k \in [K]} \pi_{kt}\ s_{j|kt} \tag{2}$$

[3]The results can easily be generalized to accommodate multidimensional endogenous variables, though we focus here on the common practice of a single endogenous variable.

[4]The model is also able to include alternative-type specific coefficients where $x_{jt}^\top = (\mathbb{1}\{j = 1\}x_t^\top, \cdots, \mathbb{1}\{j = J\}x_t^\top)$, and $\beta_\theta = \{\beta_{\theta 1}, \cdots, \beta_{\theta J}\}$. We leave the identification of such a problem into extensions.

[5]The fixed effects might be redundant by replacing the alternative-type specific fixed effects $u_{\theta j} \in U$ by setting the dummy variables as observed characteristics $\sum_{i \in [J]} u_{\theta j} \mathbb{1}\{j = i\}$ but we include the fixed effects here since the identification strategies for the alternative specific coefficients are different.

[6]Our model also extend to $v_{n\theta jt} = x_{jt}^\top \beta_\theta - \alpha_\theta p_{jt} + \lambda_\theta \xi_{jt} + e_{\theta jt} + \epsilon_{n\theta jt}$ with $e_{\theta jt}$ independent over types.

where $\pi_{kt} := \mathbb{P}(\theta = k \mid \mathcal{X}_t)$ and $s_{j|kt} := \frac{\mathbb{1}_{J_k}(j) e^{\delta_{kjt}}}{\sum_{i\in J_k} e^{\delta_{kit}}}$.[7] The indicator function, $\mathbb{1}_{J_k}(j)$, takes the value 1 if and only if $j \in J_k$. We refer to this model for aggregate market data as a Finite Mixture of Logit Models (FMML).

We will show that the FMML models simplify the estimation process by allowing for a straightforward interpretation of choice probabilities and facilitating the incorporation of observed and unobserved factors affecting the decision-making process. Due to several examples below, such a semi-parametric simplification is still practical and useful. [8]

Challenges for Identification: Identifying the parameters of model (1) nonparametrically poses significant challenges. First, the categorization of alternatives into nests is not directly observable, preventing a straightforward mapping between utilities and market shares, as in Berry (1994). Even when nests are known, the unobserved random vectors $\xi_{\theta t}$ complicate identification. In the Appendix I.2, we provide a constructive example illustrating this issue. Some models simplify the identification by ignoring unobserved product-specific characteristics (Abaluck and Adams-Prassl, 2021; Aguiar and Kashaev, 2019), while recent works incorporate these unobserved characteristics (Abaluck et al., 2020; Agarwal and Somaini, 2022). For example, Abaluck et al. (2020) use a one-to-one mapping and the connectivity assumption from Berry and Haile (2014), where changes in one product's utility affect all others, while Agarwal and Somaini (2022) assume two sets of instruments: one affecting the endogenous variable and another influencing consideration sets.

We address these unobserved characteristics with different identification assumptions and strategies. A key assumption we employ is the specific choice probability of the logit format conditional on type, which aids in distinguishing whether observed variation originates from preferences or the consideration set. This additional assumption enhances our ability to disentangle these sources of variation with only one set of instruments, thus advancing the identification strategy for the model.

[7] We have

$$\begin{aligned}\mathbb{P}(\text{argmax}_{i\in J_k} v_{n\theta it} = j \mid \theta = k, \delta_{\theta jt}) &= \mathbb{P}(\text{argmax}_{i\in J_k} v_{nkit} = j \mid \theta = k, \delta_{kjt}) \\ &= \mathbb{P}(\delta_{kjt} + \epsilon_{nkjt} > \delta_{kit} + \epsilon_{nkit}\ \forall i \neq j \mid \theta = k, \delta_{kjt}) \\ &= \frac{e^{\delta_{kjt}}}{\sum_{i\in J_k} e^{\delta_{kit}}}, \quad \forall j \in J_K.\end{aligned}$$

[8] Note that we don't consider outside options here. Still, similar to the nested logit model (Miller et al., 2021), we can consider the outside option as belonging to a distinct type 0 nest J_0, solely comprising outside options. For the outside option $s_{n0t} = 1$ if $\theta = 0$ and $s_{n0t} = 0$ if $\theta \neq 0$, we have $\mathbb{E}\left(s_{njt} \mid \theta \neq 0, \mathcal{X}_t\right) = \mathbb{E}\left[\mathbb{E}\left(s_{njt} \mid \theta, \theta \neq 0, \mathcal{X}_t\right) \mid \theta \neq 0, \mathcal{X}_t\right] = \sum_{k\in[K]} \mathbb{P}_t(\theta = k \mid \theta \neq 0) \mathbb{P}\left(\text{argmax}_{i\in J_\theta} v_{n\theta it} = j \mid \theta = k, \delta_{\theta t}\right) = \frac{s_{jt}}{1-s_{0t}}$.

2.3 Examples

This subsection provides two primary examples closely related to our approach for identification, estimation, and empirical applications.[9]

Example 1 (Nested Logit) A logit model has the assumption of independence of irrelevant alternatives (IIA). In detail, the proportions of the shares of two alternatives only depend on their characteristics (Train, 2009). As a result, logit models exhibit substitution patterns where an increase in the price of one alternative may not increase the shares of the most similar item significantly.

The nested logit addresses some of these issues. By adjusting the distribution of the error terms. Consumers select a nest with similar alternatives first and then alternatives within the nest. Partition J as $\{J_1, \ldots, J_K\}$. Let $\theta : [J] \to [K]$ be the function that assigns each alternative j to the index of the partition that contains it; that is, $j \in J_k$ if $\theta(j) = k$. The share formula becomes

$$s_{jt} = \frac{(\sum_{i \in J_{\theta(j)}} e^{\frac{\tilde{\delta}_{it}}{\sigma^{\theta(j)}}})^{\sigma^{\theta(j)}}}{\sum_{l=1}^{K}(\sum_{i \in J_l} e^{\frac{\tilde{\delta}_{it}}{\sigma^l}})^{\sigma^l}} \frac{e^{\frac{\tilde{\delta}_{jt}}{\sigma^{\theta(j)}}}}{\sum_{i \in J_{\theta(j)}} e^{\frac{\tilde{\delta}_{it}}{\sigma^{\theta(j)}}}} = \pi_{\theta(j)t} s_{j|\theta(j)}$$

where $\tilde{\delta}_{jt} = x_{jt}\beta + u_j + \xi_{jt}$ with $\sigma^k < 1$. The nested logit, therefore, is a special case of (2), and our model has more flexibility on the coefficients and allows overlapped nests to exist. Moreover, our specification admits our model to fit into the framework of the generalized nested logit proposed in Wen and Koppelman (2001) by setting the fixed effects properly. Most importantly, we do not require the "nests", and their numbers are known.

Example 2 (Logit with Limited Consideration) Limited consideration models (Abaluck and Adams-Prassl, 2021) assume

$$s_j(p_t) = \sum_k \pi_k(p_t) s_j^*(p_t \mid C_k), \tag{3}$$

where $s_j(p_t)$ is the observed probability of good j being bought given market prices p_t. $\pi_k(p_t)$ gives the probability that the set of goods C_k is considered given observable characteristics. $s_j^*(p_t|C_k)$ gives the probability that good j is chosen from the consideration set C_k. If the $s_j^*(p_t|C_k)$ satisfies a logit model, we will demonstrate that it aligns with our framework. The case is unique in terms of the identification compared to Berry (1994) in the sense that in reality, if certain

[9] Although the FMML method extends to the topic model with covariates (as discussed in Appendix H), this model's identification lies outside the paper's current scope.

product types are not considered, the share of some products might be zero.

3 Identification

This section establishes the identification of model (2) under two scenarios. First, when the nests of types are disjoint, meaning $J_k \cap J_{k'} = \varnothing$ for any $k \neq k'$. Second, when overlapping nests occur—i.e., if there exists $k \neq k'$ such that $J_k \cap J_{k'} \neq \varnothing$—, but any pair of types for which there is an overlap exhibits different price sensitivity; i.e., $\alpha_k \neq \alpha_{k'}$. If either of these scenarios holds, we say that our model has *separating* types. The primary identification assumption in this section excludes the case where a single nest contains all alternatives. The treatment of such a nest, as well as overlapping nests with possibly identical price coefficients (and thus the formal discussion of separating types), is deferred to Section 6.

3.1 Identification assumptions

Definition 1. *An alternative $j \in [J]$ is said to be a pure alternative for type $k \in [K]$ if (a) $\mathbb{1}_{J_k}(j) = 1$ and (b) $\mathbb{1}_{J_{k'}}(j) = 0$ for any $k' \neq k$.*

The main assumption we use to identify the FMML model in Equation (2) is the following:

Assumption 1. *(Pure Alternatives) For any type $k \in [K]$, at least two pure alternatives exist for each type k.*

In words, the pure alternatives assumption requires that for each type there exist at least two alternatives in his/her choice set that are not considered by any other type k. Our assumption is an extension of the "separability" condition used in the literature studying nonnegative matrix factorization problems; for example, see Theorem 4.37 in Gillis (2020). To the best of our knowledge, analogs of the separability condition have not been used to identify the type of models considered in this paper, which feature unobserved heterogeneity ξ_{jt}.[10]

We also assume there is a vector of instrumental variables $z_t = (z_{1t}, \ldots, z_{Jt})$ with $z_{jt} \in \mathbb{R}$ for the prices. For simplicity of the argument, we condition on an arbitrary value of $x_t \in \mathbb{R}^{(d-1)\times J}$ and suppress it to fixed effects with type-specific fixed effect as Berry and Haile (2014) to $u \in U^{K\times J}$. The following assumptions apply to the instrumental variables.

Assumption 2. *(Instrumental Variable) (a) For all $j \in [J]$, $\mathbb{E}(\xi_{jt} \mid z_t) = 0$. (b) For all $j, j' \in [J]$, $\mathbb{E}(p_{jt} - p_{j't} \mid z_t)$ is not a constant almost surely.*

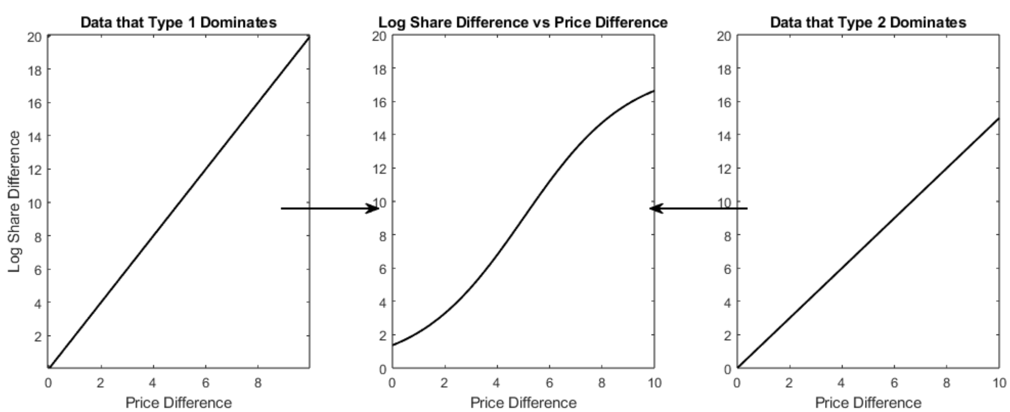
[10]The closest analog of the separability assumption in the context of models with limited consideration appears in Aguiar and Kashaev (2019), but in their framework, there are no unobservable characteristics and need only one alternative per type

We mainly consider cost shifters; for example, Petrin (2002) uses plant closures and entry costs to handle price endogeneity in the demand estimation, while other studies have relied on various supply-side instruments; for instance, cost-shifting variables such as fuel prices and emission standards, as employed in Train and Winston (2007), and tariffs and exchange rates, as demonstrated in Goldberg (1995), provide valid instruments by shifting costs across models without directly affecting consumer preferences, and may be applicable in this context by interacting with an alternative-specific dummy variable.[11]

Finally, we impose an extra condition to exploit Assumption 1 for identification. For any alternative $j \in [J]$, define $\theta(j) := \{\theta \in [K] : \mathbb{1}_{J_\theta}(j) > 0\}$. That is, $\theta(j)$ collects all types with alternative j in their consideration set. Define also the function $I(\cdot) : [K] \to 2^{[J]}$ that assigns to each type its pure alternatives. That is, $I(\theta)$ denotes all the alternatives only considered by type θ. With the definition, the collection of sets of pure alternatives is given by $\mathcal{I}_0 := \{I(1), \cdots, I(K)\}$ and the sets of all pure alternatives is given by $I_0 := \cup_{k=1}^{K} I(k)$. For two alternatives j and j', define the log share difference as

$$\Delta^{jj'} s_t := \log s_{jt} - \log s_{j't}.$$

Figure 1: Sufficient condition for the dependence of irrelevant alternatives



Assumption 3. *If alternatives j and j' are not pure alternatives of the same type, then*

$$\mathbb{E}(\Delta^{jj'} s_t \mid z_t) \neq u + a\mathbb{E}(p_{jt} - p_{j't} \mid z_t)$$

[11]We might use BLP-type product-level instrumental variables, where Assumption 2(b) requires that the instruments affect price differences between alternatives. For example, if a firm produces two vehicles with differing horsepower, a decrease in this difference should lead to a narrower price gap as the firm adjusts prices to align with relative product characteristics. However, identification warrants further discussion due to the non-fixed nature of covariates.

for any constant a and u with positive probability.

In the two remarks below, we provide intuitively sufficient conditions to support the assumption of "dependence on irrelevant alternatives," which here, to be more precise, is that the log share difference does not satisfy a linear format. We define

$$\Delta^{jj'|k} s_t := \log s_{j|kt} - \log s_{j'|kt}$$

for any $k \in \theta(j) = \theta(j')$.

Remark 3.1. *A sufficient condition ensures that Assumption 3 holds. Specifically, for any pair of alternatives j and j' that are not pure alternatives of the same type, and for any type $k \in \theta(j) \cup \theta(j')$, there exists a sequence $\{\tilde{z}_t^{k,jj'}\}$ such that a type of shares for the pair dominant:*

$$\frac{s_{jt}}{s_{j't}} \mid \tilde{z}_t^{k,jj'} \to \frac{s_{j|kt}}{s_{j'|kt}} \mid \tilde{z}_t^{k,jj'}$$

uniformly as $t \to \infty$, and $\mathbb{E}(p_{jt} - p_{j't} \mid \tilde{z}_t^{k,jj'})$ remains bounded as t grows. See the proof in A.1.

The first economic intuition behind this assumption is that, without loss of generality, for all $k' \in \theta(j) \cup \theta(j')$ with $k' \neq k$, we use instrumental variables to drive the utilities of pure alternatives $j_{k'} \neq j$ and $j_{k'} \neq j'$ to infinity, while holding the other utilities "fixed", such that $s_{j|k'} \mid \tilde{z}_t^{k,jj'} \to 0$ and $s_{j'|k'} \mid \tilde{z}_t^{k,jj'} \to 0$. In this scenario, the only attractive alternative within type k' is the pure alternative $j_{k'}$ so j and j' are ignored by consumers. The second economic intuition is that the instrument shifts π_{kt} such that $\pi_{kt} = 1$ almost surely, implying that $\pi_{k't} = 0$ for all $k' \neq k$.

Remark 3.2. *The intuitions above apply when instruments only shift preferences within a nest or alter the probability of types. The sufficient assumption in Remark 3.1 for Assumption 3 is common in many other setups, such as the nested logit model with $p_t \perp \xi_t$ for the simplicity of the argument. The proof is provided in Appendix A.1. We require this assumption because a problem arises when, in a nested logit model, the σ^k parameters are set to 1, making the model equivalent to a simple Logit model. In this case, neither the number nor the distribution of mixtures can be identified. Appendix A.1 excludes this scenario and only the pure alternatives within the same nest satisfy the IIA assumption.*

3.2 Identification of the parameters

With a known distribution of $\mathcal{S}_t := \{s_t, p_t, x_t, z_t\}$, our goal is to identify $\{\theta(\cdot), \sigma^{-1}(\cdot), \pi(\cdot), \alpha^K, u^{I_0}\}$ where the commonly used notation $\sigma^{-1}(\cdot)$: $\mathcal{S}_t \to \xi_t$ recovering the unobserved qualities and the function $\pi(\cdot)$: $\mathcal{S}_t \to \pi_t$ obtaining the probabilities of nests. $\alpha^K = (\alpha_1, \ldots, \alpha_K)$ is the price coefficients of different types. $u^{I_0} \in \mathbb{R}^{J\times K}$ represents fixed effects relative to a set of pure alternatives I_0 normalized to zero. As in the previous subsection, we condition on a given x_t.[12]

With Assumptions in 3.1, we can identify the price coefficients and the number of types. For two pure alternatives j_k and $j'_k \in I(k)$, we have that

$$\mathbb{E}(\Delta^{j_k j'_k} s_t|z_t) = u + a\mathbb{E}(p_{j_k t} - p_{j'_k t}|z_t), \tag{4}$$

where $u = u^{I_0}_{j_k} - u^{I_0}_{j'_k}$ and $a = \alpha_k$. Equation (4) is also sufficient for pure alternatives of the same type due to Assumption 3. Because j_k and j'_k are pure alternatives for type k, then $\theta(j_k) = \theta(j'_k) = k$ define the type. The following theorem formalizes the argument.

Theorem 1. *Under assumptions 1, 2, and 3, the parameter α^K in the FMML is identified.*

See the proof of Theorem 1 in Appendix A.2.

We can finish the identification directly if the nests do not overlap. As for overlapping nests, we simplify notation by defining $e^{I_0}_{kj} := \mathbb{1}\{j \in J_k\} e^{u^{I_0}_{kj}}$ for $k \in [K]$ and $j \in [J]$. We refer to $e^{I_0}_{kj}$ as the type-alternative fixed effect of a share. We want to identify these parameters because they characterize the vectors of fixed effects u^{I_0} and types of alternatives $\theta(\cdot)$. To formulate moment conditions for estimating these fixed effects, we specify the inverse function for $\tilde{\xi}_{jt}$ as:

$$\sigma^{-1}(\tilde{s}_{jt}, \tilde{s}^{I_0}_t, p_{jt}, p^{I_0}_t; \tilde{\alpha}^K, \lambda^K, \tilde{e}_{\bullet j}, \tilde{\xi}^K_t) = \tilde{\xi}_{jt}$$

where $\tilde{s}^{I_0}_t = (\tilde{s}_{j_1 t}, \cdots, \tilde{s}_{j_K t})$, $p^{I_0}_t = (p_{j_1 t}, \cdots, p_{j_K t})$, $\tilde{\alpha}^K = (\tilde{\alpha}_1, \cdots, \tilde{\alpha}_K)$, $\lambda^K = (\lambda_1, \cdots, \lambda_K)$, $\tilde{e}_{\bullet j} = (\tilde{e}_{1j}, \cdots, \tilde{e}_{Kj})$ for $j \in [J]$, $\tilde{\xi}^K_t = (\tilde{\xi}_{j_1 t}, \cdots, \tilde{\xi}_{j_K t})$, and σ^{-1} is defined through the share equation:

$$\tilde{s}_{jt} = \sum_k \frac{\tilde{e}_{kj} \tilde{s}_{j_k t} e^{\lambda_k \tilde{\xi}_{jt} - \alpha_k p_{jt}}}{e^{\lambda_k \tilde{\xi}_{j_k t} - \alpha_k p_{j_k t}}}. \tag{5}$$

The inverse function exists due to the monotonicity of equation (5). We can show that with extra assumption, $e^{I_0}_{\bullet j}$ is unique such that $\mathbb{E}[\sigma^{-1}(s_{jt}, s^{I_0}_t, p_{jt}, p^{I_0}_t; \alpha^K, \lambda^K, e^{I_0}_{\bullet j}, \xi^{I_0}_t) \mid z_t] = 0$. Therefore, parameters $\{\theta(\cdot), u\}$ are identified.

Assumption 4. *(Full Rank) For any non-negative vectors $e^K, \tilde{e}^K \in \mathbb{R}^K_+$, random vectors $\zeta^K_t, \tilde{\zeta}^K_t \in \mathbb{R}^K_+$ such that $\mathbb{E}(\zeta_t \mid z_t) = 0$ and $\mathbb{E}(\tilde{\zeta}_t \mid z_t) = 0$, if $e \neq \tilde{e}$, then*

$$\sum_{k\in[K]} e_k e^{\zeta_{kt}} \pi^*_{kjt} \neq \sum_{k\in[K]} \tilde{e}_k e^{\tilde{\zeta}_{kt}} \pi^*_{kjt}$$

[12] We can identify β once u is identified

with positive probability, where $\pi^*_{kjt} = \frac{\pi_{kt} e^{-\alpha_k p_{jt}}}{\sum_{i \in J_k} e^{\delta_{kit}}}$.

Remark 3.3. *We can rely on a similar sufficient assumption as Remark 3.1. For type* $k \in [K]$*, there exists a sequence of* $\{\tilde{z}_t^{k,j}\}$ *such that*

$$\pi^*_{kjt} \mid \tilde{z}_t^{k,j} >> \pi^*_{k'jt} \mid \tilde{z}_t^{k,j}$$

uniformly for any $k' \neq k$.

Theorem 2. *Under assumptions 1, 2, 3, and 4, the parameters* $\{\theta(\cdot), \alpha^K, u^{I_0}\}$ *in the FMML are identified.*

See the proof of Theorem 2 in Appendix A.2. If we assume the unobserved characteristics are $i.i.d.$, the variance of the pure alternatives can be used to identify λ here. This theorem suggests we can identify the pure alternatives, price coefficients, and fixed effects relative to fixed impacts of a set of pure alternatives normalized as zero by moment conditions. However, the unobserved characteristics and probabilities of nests are still interesting for the counterfactual analysis. Next, we will discuss how to identify them.

Identifying the ξ in the standard nested logit model is straightforward by normalizing one of the pure alternatives in each nest to zero, which is a common practice. However, when dealing with overlapping nests, the situation becomes more complex. Specifically, if we attempt to recenter a set of pure alternatives of different types, but some of them are "connected" by overlapping nests as zero when they are not, we introduce misspecification into the model. This misspecification affects the probabilities associated with the nests, as the characteristics of alternatives in other overlapping nests will now be included erroneously. Additionally, the error terms of these alternatives will be distorted by a similar factor, complicating the identification of ξ.

Without loss of generality, we choose a specific set of pure alternatives $I_0 = \{j_1, \ldots, j_K\}$ from different nests. We define $\xi_t^{I_0} = \{\xi_{j_1 t}, \cdots, \xi_{j_K t}\}$ as a set of intermediate parameters for the identification to reflect that they cannot be recentered as zeros. For simplicity

$$\sigma_{jt}^{-1}(e_{\bullet j}^{I_0}, \xi_t^{I_0}) := \sigma^{-1}(s_{jt}, s_t^{I_0}, p_{jt}, p_t^{I_0}; \alpha^{I_0}, \lambda, e_{\bullet j}^{I_0}, \xi_t^{I_0}).$$

With the intuition above and the intermediate parameters, the following theorem concludes the identification section.

Theorem 3. *Under assumptions 1, 2, 3, and 4, additionally assume the conditional distributions* $s_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}$ *and* $p_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}$ *are known for any* $j \in J - I_0$*, and* $E(\xi_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}) = 0$ *for any* $j \in J - I_0$*, then parameters* $\{\sigma^{-1}(\cdot), \pi(\cdot)\}$ *in the FMML are identified.*

See the detailed proof in Appendix A.2.

Remark 3.4. *Alternatively, we can also assume ξ_{jt} are i.i.d. conditional on market t, with $E(\xi_{jt}|t) = \iota_t$ where ι_t reflects market-specific information and the preference for the inside over the outside option.*[13] *We have ι_t canceled out in the share equation. Note the distinction between Theorems 2 and 3: observed variable coefficients are identified from cross-market variation, while unobserved characteristics are identified within markets.*

4 Estimation

The estimation differs from the common practice after identification since the number of observations is limited while the model is high-dimensional. Specifically, the data comprises the sales of each alternative $j \in [J]$ purchased across markets $t \in [T]$. We denote the sales and the empirical shares by the $J \times T$ matrices Y and $\hat{S}$. Let N represent the total number of consumers across all markets and the empirical share $\hat{S} = \frac{Y}{N}$.[14] For each market t, we assume that

$$Y_{\bullet t} \mid (\mathcal{X}_t, \pi_t, \theta, \alpha, \beta) \sim \text{Multinomial}(N, s_t), \tag{6}$$

where $Y_{\bullet t}$ is the t-th column of matrix Y, and $\hat{s}_t = (\hat{s}_{1t}, \ldots, \hat{s}_{Jt})$ is the t-th column of $\hat{S}$. We assume $\mathcal{X}_t$ is an independent bounded random vector for simplicity in the statistical analysis of this high-dimensional model. In the next section, we formalize the assumptions to provide statistical guarantees. One immediate concern is whether a multinomial assumption aligns with individual-level utility maximization, where $s_{njt} \mid \mathcal{X}_t$ represents a binary choice variable with expectation s_{jt}. Since the choice is mutually exclusive, the multinomial distribution is aggregate choice probabilities under individual utility maximization, where each consumer's binary choice can aggregate into a multinomial outcome.

This section presents a nested two-step algorithm for estimating the identified parameters. First, we introduce algorithms for estimating and verifying the linear functional form of log share differences between pairs of pure alternatives. Next, we estimate all remaining parameters using the identified pure alternatives through a simulated generalized method of moments (GMM) approach. Nonparametric checks for linearity are particularly challenging due to the curse of dimensionality, especially when the dimensionality of the instrumental variables is of order J. To simplify the problem, beyond assuming data independence, we impose a stronger regularity condition during the estimation process, where the instrumental variables for pairs are sufficient

[13] We omit cluster-level standard errors for simplicity, but the assumption can be extended to conditional independence between alternatives in different clusters. We leave dependence on the covariates as future works.

[14] Even if markets have varying consumer numbers, with a minimum of N, the argument remains valid.

to verify the linear specification. Formal assumptions and statistical properties are outlined and discussed in Subsection 4.2.

4.1 Algorithms

We concentrate on (4) to recover the pure alternatives first but with variation in the exogenous variable x,

$$\Delta^{jj'} s_t - (x_{jt} - x_{j't})^\top \beta - \alpha(p_{jt} - p_{j't}) - u = \lambda(\xi_{jt} - \xi_{j't}), \tag{7}$$

where $E(\xi_{jt} - \xi_{j't}|z_{jt}, z_{j't}) = 0$. The parameters $\beta = \beta_k$, $\alpha = \alpha_k$, $u = u_{kjt} - u_{kj't}$ and $\lambda = \lambda_k$ are identical for any $k \in \theta(j)$. We introduce notation for the simplicity of the arguments. Given alternatives j and j', we define $w_t^{jj'} := w_{jt} - w_{j't}$ for any random variable w_{jt}. We also introduce $\tilde{Z}^{jj'} \in \mathbb{R}^{T\times(d+2)}$ where the column vector of t-th row $\tilde{z}_t^{jj'} = (x_t^{jj'\top}, z_{jt}, z_{j't}, 1)^\top$ and $\tilde{X}^{jj'} \in \mathbb{R}^{T\times(d+1)}$ where the column vector of t-th row $\tilde{x}_t^{jj'} = (x_t^{jj'\top}, p_t^{jj'}, 1)^\top$. Alternatively, we can write

$$\Delta^{jj'} s_t - x_t^{jj'\top}\beta - \alpha p_t^{jj'} - u = \lambda \xi_t^{jj'}. \tag{8}$$

Algorithm 1 estimates the sets of pure alternatives by checking the linearity of the pairwise difference of log shares. Ideally, we rely on the conditional moment test

$$\Phi = \mathbb{E}[\tilde{\xi}_t^{jj'} \mathbb{E}(\tilde{\xi}_t^{jj'}|z_{jt}, z_{j't}) f(z_{jt}, z_{j't})],$$

and its sample analog $T^{-1}\sum_{t=1}^T \tilde{\xi}_t^{jj'} \mathbb{E}(\tilde{\xi}_t^{jj'}|z_{jt}, z_{j't}) f(z_{jt}, z_{j't})$ as Fan and Li (1996); Li and Racine (2023) to test the pure alternatives where $\tilde{\xi}_t^{jj'}$ instead of $\xi_t^{jj'}$ indicates that $\tilde{\xi}_t^{jj'} \neq \xi_{jt} - \xi_{j't}$, (j, j') might not be a pair of pure alternatives in the same nest. Specifically, $\hat{\mathcal{I}}_0$ returns a collection of sets of pure alternatives. Meanwhile, the variation in the specific product, except for the variation in the probabilities of the canceled-out types, can be pinned down by estimated pure alternatives.

Algorithm 2 estimates other parameters by the (conditional) moment conditions. For any alternative $j \in [J]$, there exists an unique vector $e_{\bullet j}^{I_0} = (e_{1j}^{I_0}, \cdots, e_{Kj}^{I_0}) > 0$ and some $\xi_t^{I_0} = (\xi_{j_1 t}^{I_0}, \cdots, \xi_{j_K t}^{I_0})$ where $\mathbb{E}(\xi_t^{I_0} \mid z_t) = 0$ such that $\mathbb{E}[\sigma^{-1}(s_{jt}, s_t^{I_0}, p_{jt}, p_t^{I_0}; \gamma^K, \lambda^K, e_{\bullet j}^{I_0}, \xi_t^{I_0}) \mid z_t] = 0$. We use the fact that ξ_{jt} are i.i.d, so given $e_{\bullet j}^{I_0}$ for all alternatives, the unobserved quality $\xi_t^{I_0}$ can be estimated by the variation within the market using $s_{jt} = \sum_k e_{kj}^{I_0} s_{j_k t} e^{\lambda_k \xi_{jt} + x_{jt}^\top \beta_k - \alpha_k p_{j_k t}} / e^{\xi_{j_k t} + x_{jt}^\top \beta_k - \alpha_k p_{j_k t}}$. We ignore the data information of alternative j for the simplicity of the notation:

$$\sigma_{jt}^{-1}(e_{\bullet j}^{I_0}, I_0, \xi_t^{I_0}) := \sigma^{-1}(s_{jt}, s_t^{I_0}, p_{jt}, p_t^{I_0}; \gamma^K, \lambda^K, e_{\bullet j}^{I_0}, \xi_t^{I_0})$$

Algorithm 1 Recover a collection of sets of pure alternatives and compute group averages

Require: Matrix of empirical shares $\hat{S} \in \mathbb{R}^{J \times T}$, matrices of exogenous covariates $X_j \in \mathbb{R}^{T \times (d-1)}$, vectors of prices $p_j \in \mathbb{R}^T$, vectors of instruments $z_j \in \mathbb{R}^T$, tuning parameter $c_0 > 0$, kernel function $k(\cdot)$, and bandwidth $h \in \mathbb{R}^2_+$

▷ for all $j \in [J]$

procedure PURE($\hat{S}, X, p, z, c_0, k, h$)

$\hat{\mathcal{I}}_0 \leftarrow \emptyset$ ▷ A collection of sets of pure alternatives

$l \leftarrow 0$ ▷ Size of $\hat{\mathcal{I}}_0$

$K \leftarrow 0$ ▷ Number of types

$\hat{\alpha}_l \leftarrow [], \hat{\beta}_l \leftarrow []$ ▷ Initialize lists to store $\hat{\alpha}_l$ and $\hat{\beta}_l$

for $j = 1, \ldots, J-1$ **do**

  for $j' = j+1, \ldots, J$ **do**

    Calculate $(\hat{\alpha}, \hat{\beta}, \hat{u}) \leftarrow$ 2SLS estimators for specification where $\hat{s}_{jt} \wedge \hat{s}_{j't} > 0$

    Calculate $\hat{\xi}_t^{jj'} \leftarrow \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > 0\} \left[\Delta^{jj'} \hat{s}_t - (x_t^{jj'})^\top \hat{\beta} - \hat{\alpha} p_t^{jj'} - \hat{u}\right]$

    Calculate $\hat{\Phi} = \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{\tau=1,\tau\neq t}^T \hat{\xi}_t^{jj'} \hat{\xi}_\tau^{jj'} h_1^{-1} h_2^{-1} k\left(\frac{z_{jt}-z_{j\tau}}{h_1}\right) k\left(\frac{z_{j't}-z_{j'\tau}}{h_2}\right)$

    if $\hat{\Phi} < \eta$ **then**

      $l \leftarrow l + 1$

      $(\hat{\mathcal{I}}_0, K(\cdot)) \leftarrow$ MERGE($\{i, j\}, \hat{\mathcal{I}}_0, l$)

      $(\hat{\alpha}_l, \hat{\beta}_l) \leftarrow (\hat{\alpha}, \hat{\beta})$

Calculate $\hat{\alpha}^K$ **and** $\hat{\beta}^K$ ▷ Group averages for each type K

for $k = 1, \ldots, \max(K)$ **do**

  $\hat{\alpha}_k^K \leftarrow \frac{1}{|\{l:K(l)=k\}|} \sum_{l:K(l)=k} \hat{\alpha}_l$

  $\hat{\beta}_k^K \leftarrow \frac{1}{|\{l:K(l)=k\}|} \sum_{l:K(l)=k} \hat{\beta}_l$

return $\hat{\mathcal{I}}_0$, $\hat{\alpha}^K$, and $\hat{\beta}^K$

function MERGE($G, \mathcal{G}_0, l$)

  $k \leftarrow 0$

  for $g \in \mathcal{G}_0$ **do**

    $k \leftarrow k + 1$

    if $g \cap G \neq \emptyset$ **then**

      $g \leftarrow g \cup G$

      $K(l) \leftarrow k$

      break ▷ No need to check further

  if no $g \in \mathcal{G}_0$ such that $g \cap G \neq \emptyset$ **then**

    $\mathcal{G}_0 \leftarrow \mathcal{G}_0 \cup \{G\}$

  return $\mathcal{G}_0$, $K(\cdot)$

Algorithm 2 Estimate the probabilities of nests

Require: matrix of empirical shares $\hat{S} \in \mathbb{R}^{J\times T}$, matrix of covariates $X_j \in \mathbb{R}^{T\times d-1}$, vectors of prices $p_j \in \mathbb{R}^T$, vectors of instruments $z_j \in \mathbb{R}^T$, collection of sets of pure alternatives $\hat{\mathcal{I}}_0$, price coefficient $\hat{\alpha}$, and vector of coefficients $\hat{\beta}$ ▷ for all $j \in [J]$
  procedure FixedEffect($\hat{S}, X, \hat{\mathcal{I}}_0, K$)
  $(\hat{I}_0, K) = \text{Combinations}(\hat{\mathcal{I}}_0)$ ▷ Sets of K alternatives from K sets in $\hat{\mathcal{I}}_0$
  $\{j_1, \cdots, j_K\} \leftarrow \mathcal{I}_0$
  for $j = 1, \cdots, J$; $t = 1, \cdots, T$ **do**
    $\hat{z}_{jt} \leftarrow (z_{j_1 t}, \cdots, z_{j_K t}, z_{jt})$
  $\hat{e} = \text{argmin}_{\tilde{e}} \sum_j \sum_t \sum_{\tau\neq t} \frac{1}{JT(T-1)} \hat{\sigma}^{-1}_{jt}(\tilde{e}, \hat{I}_0)\hat{z}^{\top}_{jt}\hat{z}_{j\tau}\hat{\sigma}^{-1}_{j\tau}(\tilde{e}, \hat{I}_0)$
  return $\hat{e}$
  procedure Nest($\hat{S}, X$)
  for $k = 1, \cdots, K$; $t = 1, \cdots, T$ **do**
    $\hat{\pi}_{kt} \leftarrow \hat{s}_{j_k t} \sum_{i\in J} \hat{e}_{ik} e^{\lambda_k \hat{\sigma}^{-1}_{it} + x'_{it}\hat{\beta}_k - \hat{\alpha}_k p_{it}} / e^{\lambda_k \hat{\sigma}^{-1}_{j_k t} + x^{\top}_{j_k t}\hat{\beta}_k - \hat{\alpha}_k p_{j_k t}}$.
  return $\hat{\Pi}$
  function $\hat{\sigma}^{-1}(\tilde{e}, \tilde{I})$ ▷ $\tilde{e} \in \mathbb{R}^{K\times J}$
  $\xi_{jt} \leftarrow 0$ for $\hat{s}_{jt} = 0$
  solve $\sum_{j\in\{J-\hat{I}_0\}\cup I(k)} \xi_{jt} = 0$ ▷ K equations
    $\hat{s}_{jt} = \sum_k \tilde{e}_{kj}\hat{s}_{j_k t} e^{\lambda_k \xi_{jt} + x^{\top}_{jt}\hat{\beta}_k - \hat{\alpha}_k p_{jt}} / e^{\lambda_k \xi_{j_k t} + x^{\top}_{jt}\hat{\beta}_k - \hat{\alpha}_k p_{j_k t}}$ ▷ $\hat{s}_{jt} > 0$ and $j \in J - \hat{I}_0$
  return ξ

There is an estimator $\hat{\sigma}^{-1}_{jt}(\tilde{e}_{\bullet j}, \tilde{\xi}^K_t)$ of the unique unobserved quality $\sigma^{-1}_{jt}(e^{I_0}_{\bullet j}, \xi^{I_0}_t)$. To be specific,

$$\hat{s}_{jt} = \sum_k \frac{\tilde{e}_{kj}\hat{s}_{j_k t} e^{\lambda_k \xi_{jt} + x^{\top}_{jt}\hat{\beta}_k - \hat{\alpha}_k p_{j_k t}}}{e^{\tilde{\xi}_{j_k t} + x^{\top}_{jt}\hat{\beta}_k - \hat{\alpha}_k p_{j_k t}}} \tag{9}$$

for $\xi_{jt} = \hat{\sigma}^{-1}_{jt}(\tilde{e}_{\bullet j}, \tilde{\xi}^K_t)$ when $\hat{s}_{jt} > 0$. [15] We rely on K equations that $\sum_{j\in\{J-\hat{I}_0\}\cup I(k)} \hat{\sigma}^{-1}_{jt}(\tilde{e}_{\bullet j}, \tilde{\xi}^K_t) = 0$ to get the estimates for K unknowns in $\xi^{I_0}_t$. $\xi^{I_0}_t$ is a function of $e^{I_0}_{\bullet j}$ and can be estimated by Algorithm 2. These equations motivate us to estimate other parameters, such as the probabilities of types, since

$$\pi_{kt} = \frac{e^{\lambda_k \xi_{j_k t} + x^{\top}_{j_k t}\beta_k - \alpha_k p_{j_k t}}}{s_{j_k t} \sum_{i\in J} e^{I_0}_{ki} e^{\xi_{it} + x^{\top}_{it}\beta_k - \alpha_k p_{it}}}.$$

Remarks regarding the implementation of the algorithms, including handling zero shares in two-stage least squares (2SLS), the selection of moments for testing functional forms, and the uniqueness of the estimates for unobserved characteristics, are provided in Appendix B.

[15] Without loss of generality, we assume that the pure alternatives always have shares greater than zero.

4.2 Statistical properties

The following assumptions address the selection of pure alternatives and the non-asymptotic property of the preference parameters of the observed covariates.

Assumption 5. *(Estimation of Pure Alternatives and Preference Parameters) (a) The coefficient spaces B^d, Λ, and U are compact subsets of $\mathbb{R}^d$, $\mathbb{R}$, and $\mathbb{R}$, respectively. (b) There is a constant $\psi > 0$ such that $\min_{k\in[K]t\in[T]} \pi_{kt} > \psi$. (c) The continuous and uniformly bounded random vectors $\mathcal{X}_t$ are i.i.d. over market $t \in [T]$. (d) Given any alternatives j and j', the eigenvalues of $\frac{Z^{jj'\top}Z^{jj'}}{T}$, $\frac{X^{jj'\top}X^{jj'}}{T}$ are uniformly strictly positive for any pair of alternatives j and j'. (e) For the instrumental variables, $\mathbb{E}(x_t\xi_t) = 0$ and $\mathbb{E}(z_t\xi_t) = 0$, and the singular value of $\frac{\tilde{X}^{jj\top}\tilde{Z}^{jj}}{n}$ is uniformly strictly positive for any pair of alternatives j and j'. (f) Given any alternatives j and $j^{'}$ that are not pure alternatives of the same type, there exists a positive constant c_s such that*

$$\min_{a,b,u} \mathbb{E}(\mathbb{E}[\Delta^{jj^{'}} s_t - x_t^{jj^{'}\top} b - ap_t^{jj^{'}} - u \mid z_{jt}, z_{j^{'}t}]^2) > c_s.$$

(g) As $T \to \infty$, $h_s \to 0$ $(s = 1, 2)$ and $Th_1h_2 \to \infty$. (h) $\log J = o(h_1h_2T)$, $JT \log^2 J = o(h_1h_2N)$.

We choose $h_s = 1.06\hat{\sigma}_s T^{-1/6}$ $(s = j, j')$ where $\hat{\sigma}_s$ is the sample standard deviation of the variable z_s.

Theorem 4. *(Pure Alternatives and Preference Parameters) Suppose Assumption 1 and 5 holds. Choose the tuning parameter $\eta = c_0\sqrt{\log(J \vee T)/h_1h_2T}$ for some $c_0 > 0$. With probability $1 - c_1(J \vee T)^{-c_2}$ for some c_1>0 and c_2>0 such that*

$$\hat{\mathcal{I}}_0 = \mathcal{I}_0$$

and for any $0 < \epsilon < 2$

$$\mathbb{P}(\|\hat{\gamma}^K - \gamma^K\|_2 > \epsilon) < C_1Te^{-C_2T\epsilon^2} + c_1(J \vee T)^{-c_2}.$$

for some C_1>0 and C_2>0.

Corollary 4.1. *(Consistency of Preference Parameters) Suppose Assumption 1 and 5 holds. Choose the tuning parameter $\eta = c_0\sqrt{\log(J \vee T)/h_1h_2T}$ for some $c_0 > 0$. Then $\hat{\gamma} \to_p \gamma$ as $T \to \infty$.*

We define

$$\tilde{\xi}_t^K(\tilde{e}) = \left\{\tilde{\xi}_t^K \in \Xi^K : \forall j \ \mathbb{E}\left[\sigma_{jt}^{-1}(s_{jt}, x_{jt}, p_{jt}) \mid t, s_t^{I_0}, j \in \{J - \mathcal{I}_0\} \cup I(k); \gamma^K, \lambda^K, \tilde{e}_{\bullet j}, \tilde{\xi}_t^K, I_0\right] = 0\right\}.$$

The two moments we rely on are

$$M(\tilde{e}) = \sum_j \frac{1}{J}\mathbb{E}\left[\sigma_{jt}^{-1}\left(\tilde{e}_{\bullet j}, I_0, \tilde{\xi}_t^K(\tilde{e})\right)\hat{z}_{jt}^{\top}\right]\left[\hat{z}_{jt}\sigma_{jt}^{-1}\left(\tilde{e}_{\bullet j}, I_0, \tilde{\xi}_t^K(\tilde{e})\right)\right],$$

and

$$\varsigma_k(\tilde{e}) = \sum_j \frac{1}{J}\mathbb{E}\left[\sigma_{jt}^{-1}\left(\tilde{e}_{\bullet j}, I_0, \tilde{\xi}_t^K(\tilde{e})\right) \mid t, j \notin \{J - \mathcal{I}_0\} \cup I(k)\right].$$

The second one is to recover $\tilde{\xi}_t^K$ given $\tilde{e}$ while the first one is to identify $\tilde{e}_{\bullet j}$ based on $\tilde{\xi}_t^K$.

Assumption 6. *(Fixed Effects, Unobserved Characteristics, and Type Probabilities) (a) ξ_{jt} are conditionally independent given the observed data. (b) A unique e^{I_0} satisfies $M(e^{I_0}) = 0$ and $\varsigma(e^{I_0}) = 0$. (c) $M(\cdot)$ and $\varsigma(\cdot)$ are twice continuously differentiable. (d) The Hessian matrices of $M(\cdot)$ and $\varsigma(\cdot)$ have strictly positive minimum eigenvalues.*

Theorem 5. *(Fixed Effects and Unobserved Characteristics) Under Assumptions 1, 5, and 6, choose the tuning parameter $\eta = c_0\sqrt{\log(J \vee T)/h_1h_2T}$ for some $c_0 > 0$. For any $\epsilon > 0$,*

$$\mathbb{P}(\|\hat{e} - e^{I_0}\|_2^2 > \epsilon) < C_1 e^{-C_2 T\epsilon^2} + c_1(J \vee T)^{-c_2}.$$

for some constant $c_1 > 0$, $c_2 > 0$, $C_1 > 0$ and $C_2 > 0$. And

$$\mathbb{P}(|\hat{\xi}_{jt} - \xi_{jt}| > \epsilon) < C_1 e^{-C_2 T\epsilon^2} + C_3 e^{-C_4 J\epsilon^2} + c_1(J \vee T)^{-c_2},$$

for some constants $c_1 > 0$, $c_2 > 0$, $C_1 > 0, C_2 > 0, C_3 > 0$, and $C_4 > 0$.

Theorem 6. *(Probability of Types) Under Assumptions 1, 5, and 6, Choose the tuning parameter $\eta = c_0\sqrt{\log(J \vee T)/h_1h_2T}$ for some $c_0 > 0$. for any $\epsilon > 0$,*

$$\mathbb{P}(\|\hat{\pi}_{kt} - \pi_{kt}\|_2^2 > \epsilon) < C_1 e^{-C_2 T\epsilon^2} + C_3 e^{-C_4 J\epsilon^2} + c_1(J \vee T)^{-c_2}.$$

for some constants $c_1 > 0$, $c_2 > 0$, $C_1 > 0, C_2 > 0, C_3 > 0$, and $C_4 > 0$.

Corollary 4.2. *(Consistency of Other Parameters) Suppose Assumption 1 and 5 holds. Choose the tuning parameter $\eta = c_0\sqrt{\log(J \vee T)/h_1h_2T}$ for some $c_0 > 0$. Then $\hat{e} \to_p e$ and $\hat{\Pi} \to_p \Pi$ as $J \wedge T \to \infty$.*

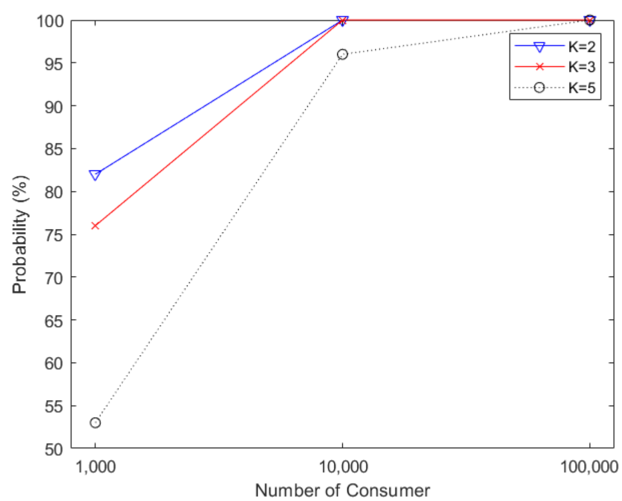
5 Empirical Evidence

5.1 Simulation

This section evaluates the effectiveness of the two-step procedure in identifying pure alternatives, estimating preference parameters, and determining the maximum distance to the fixed effects and type probabilities using randomly generated data, excluding the outside option. We consider three simulation settings: (1) the number of alternatives set to 30 and 150; (2) the number of types set to 2, 3, and 5; and (3) nested logit with inclusive parameters from 0.4 to 0.6 and FMML models with a zero substitution pattern between types.

In setting (3), both nested logit and FMML are applied to a non-overlapping structure, where price coefficients are set to -1 across all types. Additionally, we explore the FMML model with overlapping nests, where price coefficients vary from -1 to -5 for five types. Nests for types are constructed evenly across all cases. In detail, we assume that each type has two pure alternatives for overlapping nests, with all other alternatives shared across types. The lambda parameter for the Nested logit model is distributed evenly from 0.6 to 0.8. Prices for these alternatives are drawn as uniform random variables ranging from 0 to 1, varying across markets.

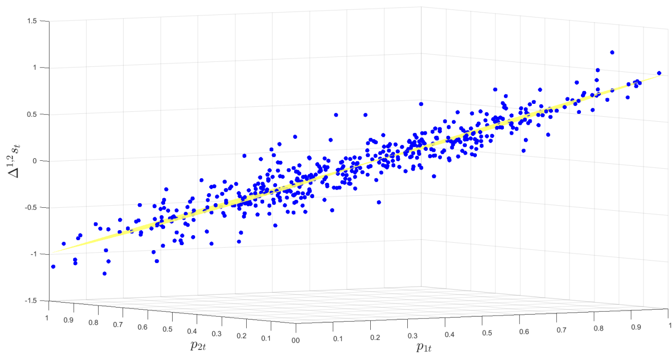
Figure 2: Predicting Probabilities in Pure Alternative Selections



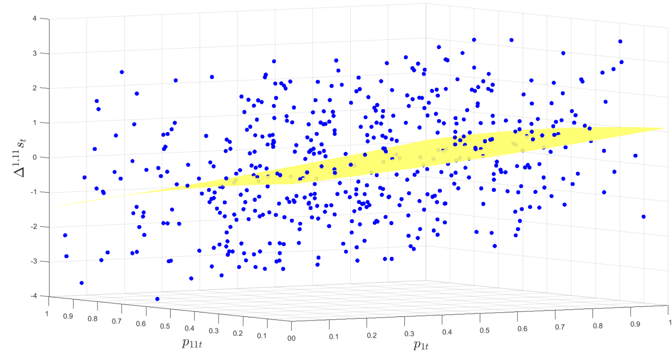
In all simulations, we set the number of consumers in each market to $N = 100{,}000$ and the number of markets to $T = 500$, as the probability of recovering the pure alternatives varies across different models from our baseline of 1,000 simulations (see Figure 2). The parameter $c_0 = 0.05$ is selected based on prediction accuracy. We use $J = 30$ alternatives, with alternative-specific fixed effects and unobserved characteristics uniformly distributed between 0 and 1, and between

0 and 0.2. Figures 3 and 6 illustrate the nested logit and FMML (without overlaps) surfaces, respectively, for $J = 30$ and $K = 3$. We observe that when a pair of alternatives belongs to the same nest, they satisfy the IIA property, whereas alternatives from different nests do not. Other data generating processes (DGPs) exhibit similar patterns.

Figure 3: Polynomial surfaces and points for same vs. different nest pairs (FMML)



(a) The same nest



(b) Different nests

Table 1: Comparison of FMML, Logit, and Nested Logit across different values of k

$J = 30$

	FMML			**Nested Logit**			**Overlapping FMML**		
	$k = 2$	$k = 3$	$k = 5$	$k = 2$	$k = 3$	$k = 5$	$k = 2$	$k = 3$	$k = 5$
FMML	1.00	1.00	1.00	1.00	1.00	1.00	(1,2)	(1,2,3)	(1,2,3,4,5)
Logit	0.97	0.91	0.85	1.50	1.50	1.58	1.46	1.90	2.72
Nested Logit	0.55	0.78	0.48	1.00	1.00	1.00			

$J = 150$

	Non-overlapping FMML			**Nested Logit**			**Non-overlapping FMML**		
	$k = 2$	$k = 3$	$k = 5$	$k = 2$	$k = 3$	$k = 5$	$k = 2$	$k = 3$	$k = 5$
FMML	1.00	1.00	1.00	1.00	1.00	1.00	(1,2)	(1,2,3)	(1,2,3,4,5)
Logit	0.99	0.92	0.99	1.43	1.49	1.56	1.46	1.92	2.79
Nested Logit	0.36	-0.12	0.27	1.00	1.00	1.00			

Table 2: Summary of second choices for EV buyers (Xing et al., 2021)

Make	Model	Fuel type	Top 1 second choice	Top 2 second choice
Honda	Accord Plug In Hybrid	PHEV	Tesla Model S	Toyota Prius
Ford	C-Max Energi	PHEV	Toyota Prius	Chevrolet Volt
Ford	Fusion Plug In Hybrid	PHEV	Chevrolet Volt	Toyota Prius Plug In
Toyota	Prius Plug-in	PHEV	Chevrolet Volt	Nissan LEAF
Chevrolet	Volt	PHEV	Toyota Prius	Nissan LEAF
Fiat	500 Electric	BEV	Nissan LEAF	Mini Cooper
Mercedes-Benz	B Class Electric	BEV	Nissan LEAF	Ford Fusion Hybrid
Ford	Focus Electric	BEV	Nissan LEAF	Chevrolet Volt
Nissan	LEAF	BEV	Chevrolet Volt	Toyota Prius
Tesla	Model S	BEV	Nissan LEAF	Audi A7
Toyota	RAV4 EV	BEV	Nissan LEAF	Tesla Model S
Chevrolet	Spark Electric	BEV	Nissan LEAF	Chevrolet Volt
Smart	fortwo electric	BEV	Nissan LEAF	Chevrolet Volt
Mitsubushi	i-MiEV	BEV	Nissan LEAF	Ford Focus Electric
BMW	i3	BEV	Nissan LEAF	Tesla Model S

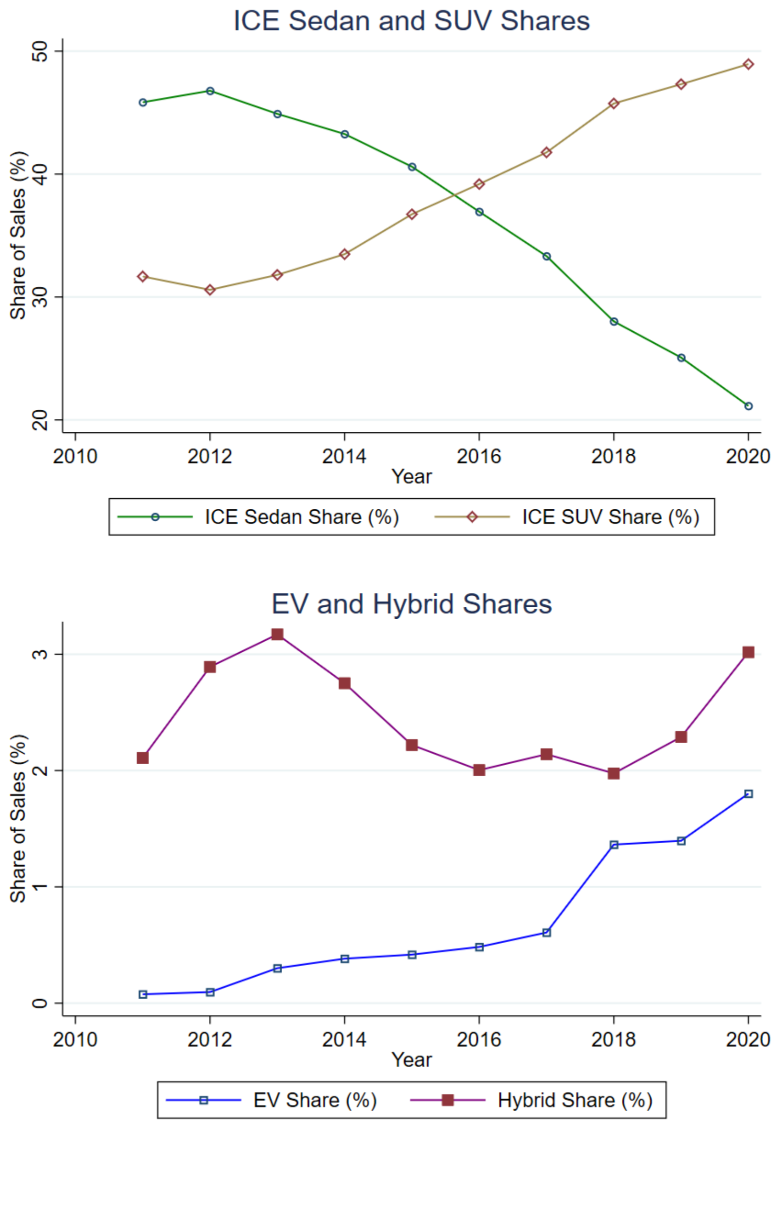
Notes: The data summary is based on the sample of 2014 EV buyers from the MaritzCX household survey data. The table summarizes the most popular alternative vehicle choices for the households that purchased different EV models. Top 1 second choice indicates the most frequently reported alternative choices among the buyers of a specific EV model. Top 2 second choice reports the second most reported alternative choices for each EV model.

Table 1 presents the preference estimates from various FMML and Nested Logit models. The columns correspond to different data-generating processes (DGPs), and the rows display the respective estimators. Based on 1,000 simulations, the FMML estimator produces unbiased results across all categories. Interestingly, the Logit estimator performs well overall. However, it tends to approximate the sample average in both the Nested Logit and Overlapping FMML models. The Nested Logit estimator is accurate only when the true underlying model is Nested Logit, even when the nest structures are correctly specified. Additionally, the simulation results show that the estimated consumer types are accurate.

5.2 Electric vehicle market

We get vehicle sales data from Wards Automotive from year 2010 to 2020. Since the panel is short, we focused on the models with the smallest test statistics. We concentrated on the two types of models, noting that results for three types yield similar conclusions. The analysis focuses on non-overlapping structures because overlapping structures produce negative price coefficients and wide tails in the estimates due to the data size. Table 2 also includes secondary choice data from

Figure 4: Annual market shares of different vehicle types in the U.S.



Xing et al. (2021), demonstrating the substitution patterns among EV consumers, particularly favoring EVs.

From the 499 models considered, we selected the top 50 models according to their shares, which existed for 10 years between 2011 and 2020. The pure alternatives chosen correspond to the models with the smallest test statistics, such as Nissan Titan ICE and F-Type ICE, as well as Chevrolet Bolt EV and Sonata PHV. Therefore, we classify the discrete into two types without overlapping nests due to the limitation of the data set. We employ a BLP-type IV, where the only IV is the average price of all other products, excluding the pair of alternatives. We then use linear regression

$$\hat{\pi}_{1t} = \beta_0 + \beta_1 \sum_{i \in J_1} e^{\hat{\delta}_{it}} + \beta_2 \sum_{i \in J_2} e^{\hat{\delta}_{it}}$$

and we find non-significant substitution patterns β_1 and β_2 between two types.

6 Extension

6.1 Multiple tiers and full consideration

The multiple-tier nested logit model extends the standard nested logit framework to accommodate more complex decision hierarchies. Alternatives are grouped into multiple levels or "tiers" of nests, reflecting a hierarchy of choices. Each tier represents a different decision-making stage, beginning with broad categories and progressing to more specific alternatives. For simplicity, consider a two-tier model:

$$s_{jt} = \sum_{k_1 \in [K_1]} \tilde{\pi}_{k_1 t} \sum_{k_2 \in [K_2]} \tilde{\pi}_{k_2 | k_1 t} \; s_{j | k_2 t}, \tag{10}$$

where $\tilde{\pi}_{k_1 t}$ is the probability of type $k_1 \in [K_1]$ in tier one, and $\tilde{\pi}_{k_2|k_1 t}$ is the conditional probability of type $k_2 \in [K_2]$ in tier two given k_1. We can then express the model as:

$$\begin{aligned}
\sum_{k_1 \in [K_1]} \tilde{\pi}_{k_1 t} \sum_{k_2 \in [K_2]} \tilde{\pi}_{k_2|k_1 t} \; s_{j|k_2 t} &= \sum_{k_1 \in [K_1]} \sum_{k_2 \in [K_2]} \tilde{\pi}_{k_1 t} \tilde{\pi}_{k_2|k_1 t} \; s_{j|k_2 t} \\
&= \sum_{k_2 \in [K_2]} \left(\sum_{k_1 \in [K_1]} \tilde{\pi}_{k_1 t} \tilde{\pi}_{k_2|k_1 t} \right) s_{j|k_2 t} \\
&= \sum_{k_2 \in [K_2]} \pi_{k_2 t} s_{j|k_2 t},
\end{aligned}$$

where $\pi_{k_2 t} := \sum_{k_1 \in [K_1]} \tilde{\pi}_{k_1 t} \tilde{\pi}_{k_2|k_1 t}$ represents the overall probability for type k_2.

An extension of this model considers a full consideration nest, where a player evaluates all alternatives. The expression that captures this is:

$$s_{jt} = \sum_{k \in [K+1]} \pi_{kt} \ s_{j|kt},$$

where $K+1$ represents the full consideration nest. By relying on pure alternatives that occur only in the full consideration nest, we can transform the problem back into our classical model. Before this transformation, we assume that $\xi_{j_{K+1}t}$ is independent of $\xi_{j'_{K+1}t}$ for other pure alternatives in the market, which allows us to identify $\xi_{J_{K+1}t}$ in the first stage. Then, observe:

$$s_{jt} - \pi_{(K+1)t} s_{j|(K+1)t} = \sum_{k \in [K]} \pi_{kt} \ s_{j|kt},$$

which enables identification of the model once $\pi_{(K+1)t} s_{j|(K+1)t}$ is known. Thereofere, the identification can work as follows: determining ξ_{jt} allows us to resolve $\pi_{(K+1)t} s_{j|(K+1)t}$, and conversely, knowing $\pi_{(K+1)t} s_{j|(K+1)t}$ enables us to identify ξ_{jt}. With mild assumptions, a one-to-one mapping exists between the market-level shares and ξ_{jt}. A formal argument for this result is left as future work.

6.2 Overlapping nests with same price coefficient

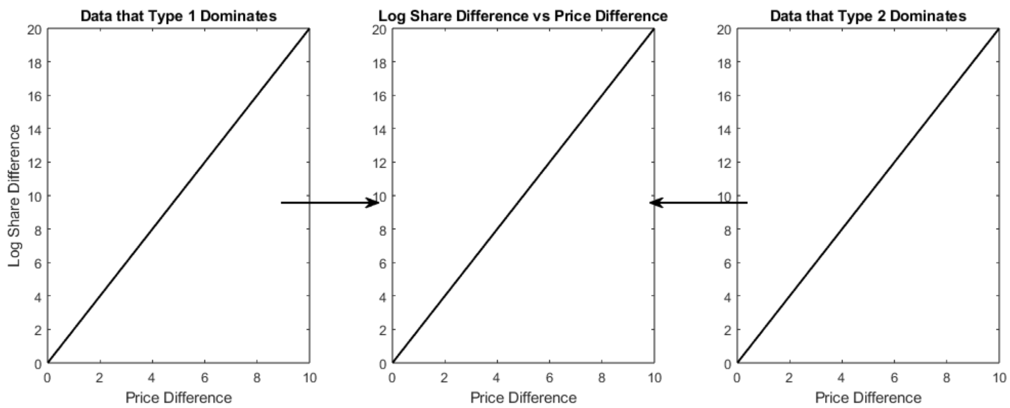
Problems would occur if different nests have the same coefficients and the nests are overlapping. For example, for alternatives j and j', where $\theta(j) = \theta(j')$, $u^{I_0}_{kj} - u^{I_0}_{kj'}$, α_k, λ_k are constants for any $k \in \theta(j)$. We still have $\mathbb{E}(\Delta^{jj'}_t s|z_t) = u + \alpha \mathbb{E}(p_{jt} - p_{j't}|z_t)$ for some constants u and a. We discuss assumptions for identifying such problematic pairs of alternatives.

Definition 2. *(Theoretically Distinct Types) Two types k and k' are said to be theoretically distinct given alternatives j and j' if*

$$\max\{|\alpha_k - \alpha_{k'}|, |(u_{kj} - u_{kj'}) - (u_{k'j} - u_{k'j'})|, |\lambda_k - \lambda_{k'}|\} > 0.$$

Definition 3. *(Separating Alternatives) Two alternatives $j \neq j'$ are said to separate types if either (a) the sets of types that consider these alternatives are different—that is, $\theta(j) \neq \theta(j')$—or (b) the sets of types that consider these alternatives are the same—that is, $\theta(j) = \theta(j')$—but there exist two types k and $k' \in \theta(j)$ that are theoretically distinct given j and j'. If j and j' are not separating alternatives, we call them non-separating alternatives.*

Figure 5: Non-separating (problematic) alternatives for the dependence of irrelevant alternatives



Assumption 7. *(Separating Condition) If alternatives j and $j^{'}$ satisfies*

$$\mathbb{E}(\Delta^{jj^{'}} s_t | z_t) = u + a\mathbb{E}(p_{jt} - p_{j^{'}t} | z_t)$$

for some constant a and u with almost surely, then j and $j^{'}$ are non-separating alternatives.

Remark 6.1. *There are multiple ways to interpret Assumption 7. For example, it reflects the difference in the "cross-price elasticity". If $\theta(j) \neq \theta(j^{'})$, without loss of generality, there is a type $k \in \theta(j)$, $k \notin \theta(j^{'})$ with a pure alternative $j_k \in I(k)$ such that $j_k \neq j$ and $j_k \neq j^{'}$.*[16] *The assumption requires that the variation of $z_{j_k t}$ on $\mathbb{E}(\log s_{jt}|z_t)$ is different with $z_{j_k t}$ on $\mathbb{E}(\log s_{j^{'}t}|z_t)$ after controlling the different effects of $z_{j_k t}$ on their prices. If $\theta(j) = \theta(j^{'})$, and the preference parameters differ, the assumption asks that the "cross-price elasticity" is different conditional on different nests.*

We know that if the alternatives are non-separating, they satisfy (4). Conversely, our Assumption 7, which is economically reasonable, guarantees the reverse implication: if (4) holds, then the pair of alternatives are non-separating. As the same as the argument in the Section 3, together with the assumption of the instrumental variable, the property of the non-separating alternatives ensures that if $\mathbb{E}(\Delta^{jj^{'}} s_t | Z = z_t) = u + a\mathbb{E}(p_{jt} - p_{j^{'}t} | z_t)$, $a = \alpha_k$ instead of any other value for any $k \in \theta(j) = \theta(j')$. Therefore, we can also obtain information on the coefficients of the characteristics by (4).

Building upon the identified non-separating alternatives and price coefficients, we develop a Vertex Hunting Method (VHM) with errors in a separate work to identify pure alternatives and their respective "fixed effects" relative to these pure alternatives. This VHM with errors extends the work of Ke et al. (2019), which does not consider unobserved quality. The VHM without

[16] The existence comes from assuming that at least two pure alternatives exist in each nest.

unobserved quality (see Arora et al. (2012a); Ke et al. (2019) and Appendix I.5) relies on the full rank condition of W. We extend it with a stronger assumption than this paper accounting for noise terms ξ. We can show that under some extra assumptions, $\mathbb{E}[\sigma^{-1}(s_{jt}, s_t^{\tilde{I}_0}, p_{jt}, p_t^{\tilde{I}_0}; \alpha^{\tilde{I}_0}, \tilde{\lambda}, \tilde{e}_{\bullet j}, \tilde{\xi}_t^{\tilde{I}_0}) \mid z_t] = 0$ for nonnegative $\tilde{e}_{\bullet j}$, nonnegative $\tilde{\lambda}$, and random variable $\tilde{\xi}_t^{I_0}$ that $E(\xi_t^{I_0}|z_t) = 0$ if and only if $\tilde{I}_0 = I_0$ for any I_0 composed by pure alternatives from different nests and $\tilde{e}_{\bullet j} = e_{kj}^{\bullet j}$.

7 Conclusion

In this paper, we have delved into the finite mixture models, emphasizing their widespread applications across various fields such as computer science, economics, and statistics. Particularly, our focus has been on Logit Models with Finite Mixtures (FMML), a category within discrete choice models, where traditional parameter estimation methods often fall short in the presence of latent subpopulations with unknown individual memberships. The paper introduces a three-step identification and estimation procedure for FMML, leveraging a sufficient assumption of pure alternatives within each mixture. This novel approach not only addresses the identification problem but also aligns with sparsity assumptions in machine learning literature, demonstrating its versatility and applicability.

Moreover, we contribute to the literature by introducing FMML as a solution for enhancing topic models in NLP and reexamining demand analysis models in the industrial organization. The possible empirical application showcases the value of FMML in providing nuanced insights that go beyond existing models' capabilities. Our method draws inspiration from machine learning techniques, departing from traditional economic approaches, such as the post-LASSO method, and emphasizing the recovery of sparse signals through clustering. This departure allows for interpretable results, aligning with the inherent sparsity in mixtures and facilitating a clearer understanding of the underlying structures of the data.

References

J. Abaluck and A. Adams-Prassl. What do consumers consider before they choose? identification from asymmetric demand responses. *The Quarterly Journal of Economics*, 136(3):1611–1663, 2021.

J. Abaluck, G. Compiani, and F. Zhang. A method to estimate discrete choice models that is robust to consumer search. Technical report, National Bureau of Economic Research, 2020.

N. Agarwal and P. J. Somaini. Demand analysis under latent choice constraints. Technical report, National Bureau of Economic Research, 2022.

V. H. Aguiar and N. Kashaev. Identification and estimation of discrete choice models with unobserved choice sets. *arXiv preprint arXiv:1907.04853*, 2019.

E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

S. Arora, R. Ge, R. Kannan, and A. Moitra. Computing a nonnegative matrix factorization–provably. In *Proceedings of the forty-fourth annual ACM Symposium on Theory of Computing*, pages 145–162, 2012a.

S. Arora, R. Ge, and A. Moitra. Learning topic models–going beyond SVD. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 1–10. IEEE, 2012b.

E. Ash and S. Hansen. Text algorithms in economics. *Annual Review of Economics*, 15:659–688, 2023.

L. Barseghyan and F. Molinari. Risk preference types, limited consideration, and welfare. *Journal of Business & Economic Statistics*, 41(4):1011–1029, 2023.

L. Barseghyan, M. Coughlin, F. Molinari, and J. C. Teitelbaum. Heterogeneous choice sets and preferences. *Econometrica*, 89(5):2015–2048, 2021a.

L. Barseghyan, F. Molinari, and M. Thirkettle. Discrete choice under risk with limited consideration. *American Economic Review*, 111(6):1972–2006, 2021b.

A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.

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

S. T. Berry and P. A. Haile. Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797, 2014.

X. Bing, F. Bunea, Y. Ning, and M. Wegkamp. Adaptive estimation in structured factor models with applications to overlapping clustering. 2020a.

X. Bing, F. Bunea, and M. Wegkamp. A fast algorithm with minimax optimal guarantees for topic models with an unknown number of topics. *Bernoulli*, 26(3):1765–1796, 2020b.

X. Bing, F. Bunea, and M. Wegkamp. Optimal estimation of sparse topic models. *The Journal of Machine Learning Research*, 21(1):7189–7233, 2020c.

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

G. Compiani and Y. Kitamura. Using mixtures in econometric models: a brief review and some new results. *The Econometrics Journal*, 19(3):C95–C127, 2016.

D. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts? *Advances in neural information processing systems*, 16, 2003.

Y. Fan and Q. Li. Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica: Journal of the econometric society*, pages 865–890, 1996.

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

N. Gillis. *Nonnegative matrix factorization*. SIAM, 2020.

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

S. Ke, J. L. M. Olea, and J. Nesbit. A robust machine learning algorithm for text analysis. Technical report, working paper, 2019.

Z. T. Ke and M. Wang. Using SVD for topic modeling. *Journal of the American Statistical Association*, pages 1–16, 2022.

D. Kozbur. Testing-based forward model selection. *American Economic Review*, 107(5):266–269, 2017.

Q. Li and J. S. Racine. *Nonparametric econometrics: theory and practice*. Princeton University Press, 2023.

W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*, 31(1):67–81, 2013.

C. F. Manski. The structure of random utility models. *Theory and decision*, 8(3):229, 1977.

X. Mao, P. Sarkar, and D. Chakrabarti. On mixed memberships and symmetric nonnegative matrix factorizations. In *International Conference on Machine Learning*, pages 2324–2333. PMLR, 2017.

D. McFadden. Conditional logit analysis of qualitative choice behavior. 1972.

G. J. McLachlan, S. X. Lee, and S. I. Rathnayake. Finite mixture models. *Annual review of statistics and its application*, 6:355–378, 2019.

N. H. Miller, G. Sheu, and M. C. Weinberg. Oligopolistic price leadership and mergers: The united states beer industry. *American Economic Review*, 111(10):3123–3159, 2021.

A. Nevo. A research assistant's guide to random coefficients discrete choice models of demand, 1998.

A. Petrin. Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, 110(4):705–729, 2002.

J. Swait and M. Ben-Akiva. Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, 21(2):91–102, 1987.

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

K. E. Train and C. Winston. Vehicle choice behavior and the declining market share of us automakers. *International economic review*, 48(4):1469–1496, 2007.

C.-H. Wen and F. S. Koppelman. The generalized nested logit model. *Transportation Research Part B: Methodological*, 35(7):627–641, 2001.

J. Xing, B. Leard, and S. Li. What does an electric vehicle replace? *Journal of Environmental Economics and Management*, 107:102432, 2021.

Contents

A Identification with unobserved characteristics

A.1 Sufficient assumption for DIA: Dominant type

Lemma A.1. *If for any pair of alternatives j and j' that are not pure alternatives of the same type, and for any type $k \in \theta(j) \cup \theta(j')$, there exists a sequence $\{\tilde{z}_t^{k,jj'}\}$ such that a type of shares for the pair dominant:*

$$\frac{s_{jt}}{s_{j't}} \mid \tilde{z}_t^{k,jj'} \to \frac{s_{j|k}}{s_{j'|k}} \mid \tilde{z}_t^{k,jj'}$$

uniformly as $t \to \infty$, and $\mathbb{E}(p_{jt} - p_{j't} \mid \tilde{z}_t^{k,jj'})$ remains bounded as t grows. Then Assumption 3 holds.

Proof. If $\theta(j) \neq \theta(j')$, to be specific, choose $k \in \theta(j)$ and $k \notin \theta(j')$, then

$$\frac{s_{jt}}{s_{j't}} \mid \tilde{z}_t^{k,jj'} \to \frac{s_{j|k}\pi_{kt}}{0} \mid \tilde{z}_t^{k,jj'} \to \infty.$$

Therefore, for any constant a, $\mathbb{E}(\Delta^{jj'} s_t \mid z_t) - a\mathbb{E}(p_{jt} - p_{j't} \mid z_t) \to \infty$. This implies that $\mathbb{E}(\Delta^{jj'} s_t \mid z_t) - a\mathbb{E}(p_{jt} - p_{j't} \mid z_t) \neq u$.

If $\theta(j) = \theta(j')$ and there exists (k_1, k_2) with $\alpha_{k_1} \neq \alpha_{k_2}$, we can choose $k = k_1$

$$\frac{s_{jt}}{s_{j't}} \mid \tilde{z}_t^{k_1,jj'} \to \frac{s_{j|k_1}}{s_{j'|k_1}} \mid \tilde{z}_t^{k_1,jj'} = e^{u_{k_1 j} - u_{k_1 j'} + \alpha_{k_1}(p_{jt} - p_{j't}) + \lambda_{k_1}(\xi_{jt} - \xi_{j't})} | \tilde{z}_t^{k_1,jj'}.$$

However, if we choose $k = k_2$,

$$\frac{s_{jt}}{s_{j't}} \mid \tilde{z}_t^{k_2,jj'} \to \frac{s_{j|k_2}}{s_{j'|k_2}} \mid \tilde{z}_t^{k_2,jj'} = e^{u_{k_2 j} - u_{k_2 j'} + \alpha_{k_2}(p_{jt} - p_{j't}) + \lambda_{k_2}(\xi_{jt} - \xi_{j't})} | \tilde{z}_t^{k_2,jj'}.$$

Taking the log and the expectation suggests a contradiction of the constant a in $\Delta\mathbb{E}(\log s_{jt} - \log s_{j't} | z_t) - a\Delta\mathbb{E}(p_{jt} - p_{j't} | z_t)$. □

Corollary A.1. *A nested logit model with $p_t \perp \xi_t$ and ξ_t bounded satisfies Assumption 3*

Proof. The construction below exactly uses the previous remark. For a nested logit model,

$$s_{jt} = \frac{e^{\frac{\delta_{jt}}{\sigma^{\theta(j)}}}}{\sum_{i \in J_k} e^{\frac{\delta_{it}}{\sigma^{\theta(j)}}}} \frac{(\sum_{i \in J_{\theta(j)}} e^{\frac{\delta_{it}}{\sigma^{\theta(j)}}})^{\sigma^{\theta(j)}}}{\sum_{l=1}^{K} (\sum_{i \in J_l} e^{\frac{\delta_{it}}{\sigma^l}})^{\sigma^l}},$$

where $\sigma^k < 1$ for $k \in [K]$ to guarantee the identification of nests.

For j and j' in different nests, set $\delta_{jt}/\sigma^{\theta(j)}$ and $\delta_{j't}/\sigma^{\theta(j')}$ to $-\infty$ with rate $-t^{\eta}$. Also, set the utility of all other elements in the nest of j' to $-\infty$ with the rate $-t^{\eta}$ and fix the utility of alternatives in $\theta(j)$.

We have

$$\begin{aligned}
\frac{s_{j't}}{s_{jt}} &= \frac{e^{\frac{\delta_{j't}}{\sigma^{\theta(j')}}}}{\sum_{i\in J_{\theta(j')}} e^{\frac{\delta_{it}}{\sigma^{\theta(j')}}}} \frac{(\sum_{i\in J_{\theta(j')}} e^{\frac{\delta_{it}}{\sigma^{\theta(j')}}})^{\sigma^{\theta(j')}}}{\sum_{l=1}^{K}(\sum_{i\in J_l} e^{\frac{\delta_{it}}{\sigma^{l}}})^{\sigma^{l}}} \Big/ \frac{e^{\frac{\delta_{jt}}{\sigma^{\theta(j)}}}}{\sum_{i\in J_{\theta(j)}} e^{\frac{\delta_{it}}{\sigma^{\theta(j)}}}} \frac{(\sum_{i\in J_{\theta(j)}} e^{\frac{\delta_{it}}{\sigma^{\theta(j)}}})^{\sigma^{\theta(j)}}}{\sum_{l=1}^{K}(\sum_{i\in J_l} e^{\frac{\delta_{it}}{\sigma^{l}}})^{\sigma^{l}}} \\
&= \frac{e^{\frac{\delta_{j't}}{\sigma^{\theta(j')}}}}{\sum_{i\in J_{\theta(j')}} e^{\frac{\delta_{it}}{\sigma^{\theta(j')}}}} (\sum_{i\in J_{\theta(j')}} e^{\frac{\delta_{it}}{\sigma^{\theta(j')}}})^{\sigma^{\theta(j')}} \Big/ \frac{e^{\frac{\delta_{jt}}{\sigma^{\theta(j)}}}}{\sum_{i\in J_{\theta(j)}} e^{\frac{\delta_{it}}{\sigma^{\theta(j)}}}} (\sum_{i\in J_{\theta(j)}} e^{\frac{\delta_{it}}{\sigma^{\theta(j)}}})^{\sigma^{\theta(j)}} \\
&\sim c_0 \frac{(\sum_{i\in J_{\theta(j)}} e^{\frac{\delta_{it}}{\sigma^{\theta(j)}}})^{1-\sigma^{\theta(j)}}}{(\sum_{i\in J_{\theta(j')}} e^{\frac{\delta_{it}}{\sigma^{\theta(j')}}})^{1-\sigma^{\theta(j')}}} \to \infty.
\end{aligned}$$

Assumption 3 reflects when all prices within a nest are very large, the nest j' itself might still be attractive. □

A.2 Price coefficient: Proof of Theorem 1

Lemma A.2. *Under Assumptions 2 and 3, alternatives j and j' are pure alternatives of the same type $k \in [K]$ with $\alpha_k = a$ if and only if there exist constants u and a such that $\mathbb{E}(\Delta_t^{jj'} s|z_t) = u + a\mathbb{E}(p_{jt} - p_{j't}|z_t)$.*

Proof. The "only if" direction is from the definition of pure alternative. We discuss the "if" direction. By Assumption 3, $\theta(j) = \theta(j')$. Suppose that (j, j') are pure alternatives. We want to show that α is unique. We have

$$\begin{aligned}
\mathbb{E}(\Delta^{jj'} s_t|z_t) &= u^{jj'} + \alpha_k\mathbb{E}(p_{jt} - p_{j't}|z_t) \\
&= u + \alpha\mathbb{E}(p_{jt} - p_{j't}|z_t).
\end{aligned}$$

Finally, $\mathbb{E}(p_{jt} - p_{j't}|z_t)$ is not degenerate and observed, so $\alpha^0 = \alpha_k$ identified. □

Theorem 1. Under assumptions 1, 2, 3, the parameter α^K in the FMML is identified.

Proof. Assumption 3 guarantees that if we observe $\mathbb{E}(\Delta^{jj'} s_t|z_t) = u + a\mathbb{E}(p_{jt} - p_{j't}|z_t)$ then the preference parameters for j and j' are the same. The following lemma proves that together with Assumption 2 if $\mathbb{E}(\Delta^{jj'} s_t|z_t) = u + a\mathbb{E}(p_{jt} - p_{j't}|z_t)$, the coefficient $\alpha_k = a$ is identified. □

A.3 Sufficient assumption for full rank: Proof of remark 3.3

Lemma A.3. *For type $k \in [K]$, there exists a sequence of $\{\tilde{z}_t^{k,j}\}$ such that*

$$\pi^*_{kjt} \mid \tilde{z}_t^{k,j} >> \pi^*_{k'jt} \mid \tilde{z}_t^{k,j}$$

uniformly for any $k' \neq k$. Then Assumption 4 is satisfied.

Proof. If

$$\sum_{k \in [K]} e_k e^{\zeta_{kt}} \pi^*_{kjt} = \sum_{k \in [K]} \tilde{e}_k e^{\tilde{\zeta}_{kt}} \pi^*_{kjt},$$

we want to show that $e = \tilde{e}$. For e_k Observe

$$\begin{aligned}
\mathbb{E}(\log \sum_{k \in [K]} e_k e^{\zeta_{kt}} \pi^*_{kjt} \mid \tilde{z}_t^{k,jj'}) &\to \mathbb{E}(\log e_k e^{\zeta_{kt}} \pi^*_{kt} \mid \tilde{z}_t^{k,jj'}) \\
&= \log e_k + \mathbb{E}(\zeta_t \mid \tilde{z}_t^{k,jj'}) + \mathbb{E}(\log \pi^*_{kt} \mid \tilde{z}_t^{k,jj'}) \\
&= \log e_k + \mathbb{E}(\log \pi^*_{kt} \mid \tilde{z}_t^{k,jj'})
\end{aligned}$$

We also have

$$\mathbb{E}(\log \sum_{k \in [K]} \pi^*_{k'jt} \mid \tilde{z}_t^{k,jj'}) \to \log \tilde{e}_k + \mathbb{E}(\log \pi^*_{kt} \mid \tilde{z}_t^{k,jj'}).$$

Therefore, $e^K = \tilde{e}^K$. □

A.4 Type-alternative specific fixed effects: Proof of Theorem 2

Lemma A.4. *Under Assumptions 2 and 4, take a set of pure alternatives in different nests $I_0 = \{j_1, \cdots, j_K\}$. For any alternative $j \in [J]$, there exists an unique vector $e^{I_0}_{\bullet j} = (e^{I_0}_{1j}, \cdots, e^{I_0}_{Kj})$ and some $\xi_t^{I_0} = (\xi_{j_1 t}, \cdots, \xi_{j_K t})$ where $\mathbb{E}(\xi_t^{I_0} \mid z_t) = 0$ such that*

$$\mathbb{E}[\sigma^{-1}(s_{jt}, s_t^{I_0}, p_{jt}, p_t^{I_0}; \alpha^K, \lambda^K, e^{I_0}_{\bullet j}, \xi_t^{I_0}) \mid z_t] = 0.$$

Proof. We start with the existence. For pure alternative j_k, by definition

$$s_{j_k t} = \pi_{kt} \frac{e^{\delta_{k j_k t}}}{\sum_{i \in J_k} e^{\delta_{kit}}} \Rightarrow \frac{\pi_{kt}}{\sum_{i \in J_k} e^{\delta_{kit}}} = \frac{s_{j_k t}}{e^{\delta_{k j_k t}}}.$$

Then, for any alternative j, we have

$$s_{jt} = \sum_{k\in[K]} \pi_{kt} \frac{\mathbb{1}_{J_k}(j)e^{\delta_{kjt}}}{\sum_{i\in J_k} e^{\delta_{kit}}} = \sum_{k\in[K]} \frac{s_{j_k t}}{e^{\delta_{kj_k t}}} \mathbb{1}_{J_k}(j)e^{\delta_{kjt}} = \sum_{k\in[K]} \frac{s_{j_k t}}{e^{-\alpha_k p_{j_k t}+\lambda_k \xi_{j_k t}}} \mathbb{1}_{J_k}(j) e^{u_{kj}^{I_0}} e^{-a_k p_{jt}+\lambda_k \xi_{jt}}$$
$$= \sum_{k\in[K]} \frac{e_{kj}^{I_0} s_{j_k t} e^{\lambda_k \xi_{jt} - a_k p_{jt}}}{e^{\lambda_k \xi_{j_k t} - \alpha_k p_{j_k t}}}.$$

By Assumption 2 we therefore have,

$$\mathbb{E}[\sigma^{-1}(s_{jt}, s_t^{I_0}, p_{jt}, p_t^{I_0}; \alpha^K, \lambda^K, e_{\bullet j}^{I_0}, \xi_t^{I_0}) \mid z_t] = 0.$$

For the uniqueness, assume that there is $\tilde{e}_{\bullet j}^K = (\tilde{e}_{1j}, \cdots, \tilde{e}^{Kj)}$ and some $\tilde{\xi}_t^K = (\tilde{\xi}_{j_1 t}, \cdots, \tilde{\xi}_{j_K t})$ such that

$$\mathbb{E}[\sigma^{-1}(s_{jt}, s_t^{I_0}, p_{jt}, p_t^{I_0}; \alpha^K, \lambda^K, \tilde{e}_{\bullet j}, \tilde{\xi}_t^K) \mid z_t] = 0,$$

which is equivalent to

$$s_{jt} = \sum_k \frac{\tilde{e}_{kj} s_{j_k t} e^{\tilde{\lambda}_k \tilde{\xi}_{jt} - a_k p_{jt}}}{e^{\tilde{\lambda}_k \tilde{\xi}_{j_k t} - \alpha_k p_{j_k t}}} = \sum_k \pi_{kt} \frac{e^{\lambda_k \xi_{j_k t} - \alpha_k p_{j_k t}}}{\sum_{i\in J_k} e^{\delta_{kit}}} \frac{\tilde{e}_{kj} e^{\tilde{\lambda}_k \tilde{\xi}_{jt} - a_k p_{jt}}}{e^{\tilde{\lambda}_k \tilde{\xi}_{j_k t} - \alpha_k p_{j_k t}}} = \sum_k \pi_{kt} \frac{e^{\lambda_k \xi_{j_k t}}}{\sum_{i\in J_k} e^{\delta_{kit}}} \frac{\tilde{e}_{kj} e^{\tilde{\lambda}_k \tilde{\xi}_{jt} - a_k p_{jt}}}{e^{\tilde{\lambda}_k \tilde{\xi}_{j_k t}}}$$
$$= \sum_k \pi_{kjt}^* \frac{e^{\lambda_k \xi_{j_k t}} \tilde{e}_{kj} e^{\tilde{\lambda}_k \tilde{\xi}_{jt}}}{e^{\tilde{\lambda}_k \tilde{\xi}_{j_k t}}} = \sum_k e^{\lambda_k \xi_{j_k t} + \tilde{\lambda}_k (\tilde{\xi}_{jt} - \tilde{\xi}_{j_k t})} \tilde{e}_{kj} \pi_{kjt}^* \qquad (\text{where } \mathbb{E}(\tilde{\xi}_t^K \mid z_t) = 0)$$

since $\pi_{kjt}^* = \frac{\pi_{kt} e^{-\alpha_k p_{jt}}}{\sum_{i\in J_k} e^{\delta_{kit}}}$ by definition.

We also know that,

$$s_{jt} = \sum_{k\in[K]} \pi_{kt} \frac{\mathbb{1}_{J_k}(j) e^{\delta_{kjt}}}{\sum_{i\in J_k} e^{\delta_{kit}}} = \sum_{k\in[K]} \pi_{kt} \frac{e_{kj}^{I_0} e^{\lambda_k \xi_{jt} - a_k p_{jt}}}{\sum_{i\in J_k} e^{\delta_{kit}}} = \sum_{k\in[K]} e^{\lambda_k \xi_{jt}} e_{kj}^{I_0} \pi_{kjt}^*$$

By assumption 4, we must have that $\tilde{e}_{\bullet j} = e_{\bullet j}$. □

Theorem 2. Under assumptions 1, 2, 3, and 4, the parameters $\{\theta(\cdot), \alpha^K, u^{I_0}\}$ in the FMML are identified.

Proof. Theorem 2 is based on the fact that the fixed effects of alternatives can be uniquely identified by a set of pure alternatives using conditional moment conditions by Lemma A.4. □

A.5 Unobserved characteristics and probabilities of types: Proof of Theorem 3

Lemma A.5. *Given ($\theta(\cdot)$, α^K, u^{I_0}), if $E(\xi_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}) = 0$ for any $j \in J - I_0$, then π_t and $\xi_t^{I_0}$ are unique for the known distributions of $s_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}$ and $p_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}$ for any $j \in J - I_0$.*

Proof. We ignore the subscript t for the simplicity of the proof since we care about the variation within the market. First, we know that for any $j \in J_k - \{j_k\}$,

$$\log s_{j|k} - \log s_{j_k} = \alpha_k(p_j - p_{j_k}) + u_{kj} - u_{j_k} + \lambda_k(\xi_j - \xi_{j_k}).$$

This implies $\mathbb{E}(\log s_{j|k} - \log s_{j_k} | \xi^{I_0}, s^{I_0}, p^{I_0}) = \alpha_k \mathbb{E}(p_j | \xi^{I_0}, s^{I_0}, p^{I_0}) - p_{j_k} + u_j - u_{j_k} + \lambda_k \xi_{j_k}$. Therefore, we can identify ξ_{j_k} by the other pure alternatives in the nests because all the other variables in the equation are observed.

With a give ξ^{I_0}, we can identify all the ξ_j. Specifically,

$$s_j = \sum_{k \in [K]} \pi_k \frac{\mathbb{1}_{J_k}(j) e^{\delta_{kj}}}{\sum_{i \in J_k} e^{\delta_{ki}}} = \sum_{k \in [K]} \frac{s_{j_k}}{e^{\xi_{j_k}}} \mathbb{1}_{J_k}(j) e^{\delta_{kj}} = \sum_{k \in [K]} \frac{s_{j_k}}{e^{\xi_{j_k}}} \mathbb{1}_{J_k}(j) e^{x_j^\top \beta_k - \alpha_k p_j + \lambda_k \xi_j}$$

by (2). There is an ξ_j by the intermediate value theorem, and the unobserved quality's exponential is monotone, so the solution is unique. □

Theorem 3. Under assumptions 1, 2, 3, and 4, additionally assume the conditional distributions $s_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}$ and $p_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}$ are known for any $j \in J - I_0$, and $E(\xi_{jt} \mid t, s_t^{I_0}, p_t^{I_0}, \xi_t^{I_0}) = 0$ for any $j \in J - I_0$, then parameters $\{\sigma^{-1}(\cdot), \pi(\cdot)\}$ in the FMML are identified.

Proof. By Theorem 2, we are able to identify $(\theta(.), u, \alpha)$. Then, conditional on each market with many products, we can identify ξ_{jt} by Lemma A.5. Then, the final step is to identify π_t, which is straightforward with ξ_{jt} identified. We plug the identified parameters into the equation of pure alternatives,

$$s_{j_k t} = \pi_{kt} \frac{\mathbb{1}_{J_k}(j) e^{\delta_{k j_k t}}}{\sum_{i \in J_k} e^{\delta_{kit}}}.$$

□

B Algorithms

B.1 Two-stage least squares

We want to explain the specific 2SLS procedure here. Before formally explaining our algorithm, the following is the oracle case if the actual s_{jt} are observed. Under the ideal case, the dependent variable is $\Delta^{jj'} s_t$. The exogenous variables are $x_t^{jj'}$ and a constant term. The endogenous variable is $p^{jj'}$ and the instruments are z_{jt} and $z_{j't}$.

Stage 1: regress endogenous variable on instruments and exogenous variables:

$$p^{jj'} = \zeta_0 + x_t^{jj'\top}\zeta_1 + \zeta_2 z_{jt} + \zeta_3 z_{j't} + e_t.$$

Obtain the fitted values from this regression:

$$\hat{p}^{jj'} = \hat{\zeta}_0 + \hat{x}_t^{jj'\top}\hat{\zeta}_1 + \hat{\zeta}_2 z_{jt} + \hat{\zeta}_3 z_{j't}.$$

Stage 2: regress the dependent variable on the fitted values and exogenous variables:

$$\Delta^{jj'} s_t = x_t^{jj'\top}\beta + \alpha\hat{p}^{jj'} + u^{jj'} + \tilde{\xi}_t.$$

The matrix format of this 2SLS estimator:

$$(\beta^{jj'}_{2SLS}; \alpha^{jj'}_{2SLS}; u^{jj'}_{2SLS}) = \left(\tilde{X}^{jj'\top}\tilde{Z}^{jj}(\tilde{Z}^{jj\top}\tilde{Z}^{jj})^{-1}\tilde{Z}^{jj\top}\tilde{X}^{jj'}\right)^{-1}\tilde{X}^{jj'\top}\tilde{Z}^{jj}(\tilde{Z}^{jj\top}\tilde{Z}^{jj})^{-1}\tilde{Z}^{jj\top}\Delta^{jj'} s.$$

The concern is that we cannot observe s_{jt}. We will show that after dropping some "zeros", the empirical shares version of the 2SLS has good non-asymptotic results close to the true parameters. In the second stage, the objective function becomes

$$(\hat{\beta}^{jj'}, \hat{\alpha}^{jj'}, \hat{u}^{jj'}) = \text{argmin}_{\alpha,\beta,u}\frac{1}{T}\sum_t \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > 0\}(\Delta^{jj'}\hat{s}_t - x^{jj'}\beta - \alpha\hat{p}^{jj'} - u)^2.$$

B.2 Functional form test

For testing the correctness of the parametric regression model of a pair of alternatives, the null hypothesis is

$$H_0^{jj'} : \mathbb{E}(\Delta^{jj'} s_t - m(\tilde{x}_t^{jj'}, \gamma) \mid \tilde{z}_t^{jj'}) = 0, \text{almost surely for some } \gamma = \gamma_0$$

where $m(\tilde{x}_t^{jj'}, \gamma)$ is a known function with γ being a vector of unknown parameters in a compact space. The alternative hypothesis is the negation of the $H_0^{jj'}$, namely,

$$H_1^{jj'} : \mathbb{E}(\Delta^{jj'} s_t - m(\tilde{x}_t^{jj'}, \gamma) \mid \tilde{z}_t^{jj'}) \neq 0, \text{with positive probability for any } \gamma.$$

The compactness of γ is required to ensure that the positive probability in the alternative hypothesis is strictly greater than zero. In our specification, we define $\Delta^{jj'} s_t = x_t^{jj'\top}\beta + \alpha p^{jj'} +$

$u^{jj'} + \tilde{\xi}_t^{jj}$ by setting $m(\tilde{x}_t, \gamma_0) = x_t^{jj'\top}\beta + \alpha p^{jj'} + u^{jj'}$. We can, therefore, base our test on the moment $\mathbb{E}[\mathbb{E}(\tilde{\xi}_t^{jj'} \mid Z_t)^2]$ to distinguish the null and the alternative.

There are two challenges to implement the estimate of $\mathbb{E}[\mathbb{E}(\tilde{\xi}_t^{jj'} \mid Z_t)^2]$ by samples. The first is that we need to estimate $f(\tilde{\xi}_t^{jj'} \mid Z_t)$. The conditional expectation, therefore, includes a term of denominator, which could approximate zero. In details, The kernel-based estimator for the conditional expectation $\mathbb{E}(u_i|X_i = x)$ is given by

$$\hat{\mathbb{E}}(u_i|X_i = x) = \frac{\sum_{i=1}^n u_i K_h(x - X_i)}{\sum_{i=1}^n K_h(x - X_i)}$$

where kernel functions are crucial in several machine learning algorithms, and they typically possess the following properties: (a)$K(x, y) \geq 0$ (b)$K(x, y) = K(y, x)$ (c) There exists a constant M such that $K(x, y) \leq M$ for all x, y. We use the Gaussian kernel, also known as the Radial Basis Function (RBF) kernel: $K(x, y) = \exp\left(-\frac{\|x-y\|_2^2}{2\sigma^2}\right)$ where σ is the kernel width parameter.

Secondly, $\mathbb{E}[\mathbb{E}(\tilde{\xi}_t^{jj'} \mid \tilde{Z}_t)^2]$ includes three expectation so three summations. We rely on the law of iterated expectation to transform $\mathbb{E}(\tilde{\xi}_t^{jj'} \mid \tilde{Z}_t)^2$ into $\mathbb{E}[\tilde{\xi}_t^{jj'}\mathbb{E}(\tilde{\xi}_t^{jj'} \mid \tilde{Z}_t) \mid \tilde{Z}_t]$ because $\mathbb{E}(\tilde{\xi}_t^{jj'} \mid \tilde{Z}_t)$ is a function of $\tilde{Z}_t$. Then we use the law of iterated expectation again $\mathbb{E}[\tilde{\xi}_t^{jj'}\mathbb{E}(\tilde{\xi}_t^{jj'} \mid \tilde{Z}_t)]$.

The final moment we use for the test is

$$\Phi = \mathbb{E}[\tilde{\xi}_t^{jj'}\mathbb{E}(\tilde{\xi}_t^{jj'} \mid z_{jt}, z_{j't}) f(z_{jt}, z_{j't})].$$

which restricts $\tilde{Z}_t$ by z_{jt} and $z_{j't}$ due to the curse of dimensionality. We estimate the $\mathbb{E}(\tilde{\xi}_t^{jj'} \mid z_{jt}, z_{j't}) f(z_{jt}, z_{j't})$ by the leave-one-out kernel estimator $\frac{1}{T-1}\sum_{\tau\neq t}\hat{\xi}_t^{jj'} K_{h,t\tau}$ where $\hat{\xi}_t^{jj'} = \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \delta\}(\Delta^{jj'}\hat{s}_t - x_t^{jj\top}\hat{\beta}^{jj'} - \hat{\alpha}^{jj'} p_t^{jj'} - \hat{u}^{jj'})$ and $K_{h,t\tau} = h_1^{-1}h_2^{-1}k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2})$. Similar to the discussion of the 2SLS, the sample analogue becomes

$$\hat{\Phi}^{jj'} = \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,z\neq t}^{T}\hat{\xi}_t^{jj'}\hat{\xi}_\tau^{jj'} h_1^{-1}h_2^{-1}k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2}).$$

B.3 Conditional GMM for fixed effects

The functional form test is designed to estimate the preference parameters. We discuss the GMM to estimate the fixed effects. First, given the alternatives' fixed effects and observed preferences, we would like to recover the unobserved characteristics. We assume the error terms are independent, allowing us to treat the unobserved characteristics of a set of pure alternatives across different nests as parameters. We use the moment condition that

$$\mathbb{E}[\sigma_{jt}^{-1}(s_{jt}, x_{jt}, p_{jt}) \mid t, s_t^{I_0}, j \notin I_0; \gamma^K, \lambda^K, e_{\bullet j}^{I_0}, \xi_t^{I_0}, I_0] = 0$$

Then, we can define

$$\tilde{\xi}_t^K(\tilde{\gamma}^K, \tilde{e}_{\bullet j}) = \{\tilde{\xi}_t^K \in \Xi^K : \mathbb{E}[\sigma_{jt}^{-1}(s_{jt}, x_{jt}, p_{jt}) \mid t, s_t^{I_0}, j \notin J - I(k); \tilde{\gamma}, \lambda, \tilde{e}_{\bullet j}, \tilde{\xi}_t^K, \tilde{I}_0] = 0\}$$

and $\tilde{\xi}_t^K = \xi_t^{I_0}$ when $\tilde{\gamma}^K = \gamma^K$ and $\tilde{e}_{\bullet j} = e_{\bullet j}^{I_0}$. Once adjusting the identified observed and unobserved qualities, we can use the shares of the pure alternatives to predict the remaining shares perfectly. These two components interact and are combined into the moment conditions:

$$\mathbb{E}[\hat{z}_{jt}\sigma_{jt}^{-1}\left(\gamma^K, \lambda, e_{\bullet j}^{I_0}, \tilde{\xi}_t(\gamma^K, e_{\bullet j}^{I_0})\right)] = 0$$

It is easy to show that the estimate is unique by the fixed point theorem. Then, we use the leave-one-out estimator for the moment condition:

$$\min_{e \geq 0} \sum_{j \in J} \frac{1}{T(T-1)} \sum_{t} \sum_{\tau \neq t} \hat{\xi}_{jt}\hat{z}_{jt}^{\top}\hat{z}_{j\tau}\hat{\xi}_{j\tau}.$$

C Non-asymptotic properties for the two stage least square

We proceed to discuss the 2SLS estimates and demonstrate that the estimate is close to the Oracle 2SLS. With an accurate estimate, we can then apply the functional form test to check for linearity. Our approach differs from the existing literature in that we examine all results from a non-asymptotic perspective. A key assumption is that the utilities of the alternatives are bounded, ensuring that the shares remain sufficiently large with high probability. Under this assumption, the logarithmic transformation will capture certain linear properties. Here, we show that if we can observe the actual shares and pure (non-separating) alternatives, we can estimate the preference parameter "consistently". Without further explanation, we might ignore the subscript of jj' for the simplicity of the argument.

Lemma C.1. *Given any non-separating alternatives j and j', define $\gamma_{2SLS}^u = (\beta_{2SLS}, \alpha_{2SLS}, u_{2SLS})$ as estimates of the Oracle 2SLS procedure given in Appendix B. If Assumption 5 holds, then there is $\gamma_k^u := (\beta_k, \alpha_k, u_k^{jj'})$ for $k \in [K]$ such that for any $\epsilon > 0$*

$$\mathbb{P}(\|\gamma_{2sls}^u - \gamma_k^u\|_2 > \epsilon) < C_1 e^{-C_2 T \epsilon^2}$$

with positive constants C_1 and C_2.

Proof. To begin with, we know that for non-separating alternatives j and $j^{'}$, for any $k \in \theta(j)$

$$\Delta^{jj^{'}} s_t = x_t^{jj^{'}\top}\beta_k - \alpha_k p_t^{jj^{'}} + u_k^{jj^{'}} + \lambda_k \xi_t^{jj^{'}}.$$

This is equivalent to

$$\Delta^{jj^{'}} s = \tilde{X}^{jj^{'}}\gamma^u + \lambda_k \xi_t^{jj^{'}}.$$

We ignore the subscript of $jj^{'}$ for the simplicity of the argument. The closed form 2SLS estimator, therefore, is,

$$\gamma^u_{2sls} = \left(\tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top \tilde{X}\right)^{-1} \tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top \Delta s.$$

As the common practice for the 2SLS, we have

$$\gamma^u_{2sls} = \left(\tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top \tilde{X}\right)^{-1} \tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top (\tilde{X}\gamma^u + \lambda_k \xi_t)$$

and therefore,

$$\gamma^u_{2sls} - \gamma^u = \lambda_k \left(\tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top \tilde{X}\right)^{-1} \tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top \xi_t.$$

By the CS inequality,

$$\begin{aligned}
\|\gamma^u_{2sls} - \gamma^u\|_2 &\leq \lambda_k \|(\frac{1}{T}\tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top \tilde{X})^{-1}\|_2 \|\frac{1}{T}\tilde{X}^\top \tilde{Z}\|_2 \|(\frac{1}{T}\tilde{Z}^\top \tilde{Z})^{-1}\|_2 \|\frac{1}{T}\tilde{Z}^\top \xi\|_2 \\
&\leq \tilde{C}_1 \|\frac{1}{T}\tilde{Z}^\top \xi\|_2,
\end{aligned}$$

where the second inequality is from the fact that $\|(\frac{1}{T}\tilde{X}^\top \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1}\tilde{Z}^\top \tilde{X})^{-1}\|_2$ is bounded by Lemma F.13 and $\|\frac{1}{T}\tilde{X}^\top \tilde{Z}\|_2$ is the sample average of bounded random variables. $\|(\frac{1}{T}\tilde{Z}^\top \tilde{Z})^{-1}\|_2$ is bounded by the minimum eigenvalue of $\frac{1}{T}\tilde{Z}^\top \tilde{Z}$. Specifically, define

$$M = \frac{1}{T}\tilde{X}^{jj^{'}\top}\tilde{Z},$$

and $M \in \mathbb{R}^{(d+1)\times(d+2)}$. For each entry, we have

$$M_{ii'} = \frac{1}{T}\sum_{t=1}^{T} \tilde{X}_{ti}\tilde{Z}_{i't}$$

bounded. By the inequality of matrix norms that $\|M\|_2 \leq \|M\|_1 \|M\|_\infty$,$\|M\|_2$ is bounded. Therefore, $\tilde{C}_1$ is a positive constant.

We only need to deal with $\|\frac{1}{T}\tilde{Z}^\top \xi\|_2$, which equals to

$$\|\frac{1}{T}\sum_t \tilde{z}_t\xi_t - \mathbb{E}(\tilde{z}_t\xi_t)\|_2 \leq \sum_{i=1}^{d+2} |\frac{1}{T}\sum_t \tilde{z}_{ti}\xi_t - \mathbb{E}(\tilde{z}_t\xi_{ti})|.$$

where we use l_2 vector norm is smaller than l_1 norm.

$$\mathbb{P}(\|\gamma^u_{2sls} - \gamma^u_k\|_2 > \epsilon) \leq P(C^k \|\frac{1}{T}\tilde{Z}^\top \xi\|_2 > \epsilon) < C_1 e^{-C_2 T \epsilon^2}$$

by Hoeffding's inequality. □

Lemma C.2. *Given alternatives $(j, j') \in [J]$ with bounded utilities such that $\frac{h_0}{J} < s_{jt} < \frac{h_1}{J}$ and $\frac{h_0}{J} < s_{j't} < \frac{h_1}{J}$, define $\hat{\gamma}^u = (\hat{\beta}, \hat{\alpha}, \hat{u})$ as estimates of the 2SLS procedure given in Algorithm 1. If $J = o(N)$, $\delta \leq \frac{c_0}{2J}$ and Assumption 5 holds, then with $\gamma^u_{2sls} := (\beta_{2sls}, \alpha_{2sls}, u_{2sls})$, for any* $0 < \epsilon < 2$

$$\mathbb{P}(\|\hat{\gamma}^u - \gamma^u_{2SLS}\|_2 > \epsilon) < C_1 T (e^{-C_2 N \epsilon^2 / J \log^2 J} + e^{-C_2 N/J})$$

with positive constants C_1 and C_2.

Proof. In the following argument, we ignore jj' for the simplicity of the statement. In reality, we optimize the objective function,

$$\hat{Q}_T(\gamma^u) = \frac{1}{T}\sum_t \mathbb{1}\{\hat{s}_t \wedge \hat{s}_t > 0\}(\Delta\hat{s}_t - x_t\beta - \alpha_{2sls}\hat{p}_t - u_t)^2.$$

By contrast, under the Oracle case, ideally, we would like to optimize,

$$Q_T(\gamma^u) = \frac{1}{T}\sum_t (\Delta s_t - x_t\beta - \alpha\hat{p}_t - u_t)^2.$$

The matrix format of this objective function is

$$Q_T(\gamma^u) = \frac{1}{T}(\Delta s - \tilde{X}\gamma^u)^\top P_{\tilde{Z}}(\Delta s - \tilde{X}\gamma^u)$$

where $P_{\tilde{Z}}$ is the projection matrix onto the space spanned by the instruments $\tilde{Z}$:

$$P_{\tilde{Z}} = \tilde{Z}(\tilde{Z}^\top\tilde{Z})^{-1}\tilde{Z}^\top.$$

We concentrate on the event

$$\{Q_T(\hat{\gamma}^u) - Q_T(\gamma^u_{2sls}) > \epsilon\}$$

because it can bound $\hat{\gamma}^u - \gamma^u_{2sls}$.

By definition of the minimization problem, we notice $\hat{Q}_T(\hat{\gamma}^u) \leq \hat{Q}_T(\gamma^u_{2sls})$ and $Q_T(\gamma^u_{2sls}) \leq Q_T(\hat{\gamma}^u)$. Since

$$Q_T(\hat{\gamma}^u) - Q_T(\hat{\gamma}^u) + \hat{Q}_T(\hat{\gamma}^u) \leq Q_T(\gamma^u_{2sls}) - Q_T(\gamma^u_{2sls}) + \hat{Q}_T(\gamma^u_{2sls}),$$

we have

$$Q_T(\hat{\gamma}^u) - Q_T(\gamma^u_{2sls}) \leq Q_T(\hat{\gamma}^u) - \hat{Q}_T(\hat{\gamma}^u) + \hat{Q}_T(\gamma^u_{2sls}) - Q_T(\gamma^u_{2sls}).$$

Moreover, since $Q_T(\hat{\gamma}^u) - Q_T(\gamma^u_{2sls}) \geq 0$, we derive the bound

$$|Q_T(\hat{\gamma}^u) - Q_T(\gamma^u_{2sls})| \leq |\hat{Q}_T(\hat{\gamma}^u) - Q_T(\hat{\gamma}^u)| + |\hat{Q}_T(\gamma^u_{2sls}) - Q_T(\gamma^u_{2sls})|.$$

By Lemma F.28 for any $0 < \epsilon < 2$,

$$\mathbb{P}(\sup_{\gamma^u} |Q_T(\gamma^u) - \hat{Q}_T(\gamma^u)| > \epsilon) < \tilde{C}_1 T(e^{-\tilde{C}_2 N\epsilon^2/J\log^2 J} + e^{-\tilde{C}_2 N/J\log^2 J})$$

with positive constants $\tilde{C}_1$, and $\tilde{C}_2$. Therefore, we derive

$$\mathbb{P}(|\hat{Q}_T(\hat{\gamma}^u) - Q_T(\hat{\gamma}^u)| + |\hat{Q}_T(\gamma^u_{2sls}) - Q_T(\gamma^u_{2sls})| > \epsilon) < \bar{C}_1 T(e^{-\bar{C}_2 N\epsilon^2/J\log^2 J} + e^{-\bar{C}_2 N/J\log^2 J})$$

by the union bound. $\bar{C}_1$, and $\bar{C}_2$ are positive constants. It suggests that

$$\mathbb{P}(|Q_T(\hat{\gamma}^u) - Q_T(\gamma^u_{2sls})| > \epsilon) < \bar{C}_1 T(e^{-\bar{C}_2 N\epsilon^2/J\log^2 J} + e^{-\bar{C}_2 N/J\log^2 J}).$$

Then we can bound $\|\hat{\gamma}^u - \gamma^u_{2SLS}\|_2$ by

$$
\begin{aligned}
|Q_T(\hat{\gamma}^u) - Q_T(\gamma^u_{2SLS})| &= \frac{1}{T}|(\Delta s - \tilde{X}\hat{\gamma}^u)^\top P_{\tilde{Z}}(\Delta s - \tilde{X}\hat{\gamma}^u) - (\Delta s - \tilde{X}\gamma^u_{2SLS})^\top P_{\tilde{Z}}(\Delta s - \tilde{X}\gamma^u_{2SLS})| \\
&= \frac{1}{T}|\Delta s^\top P_{\tilde{Z}}\Delta s - 2\Delta s^\top P_{\tilde{Z}}\tilde{X}\hat{\gamma}^u + \hat{\gamma}^{u\top}\tilde{X}^\top P_{\tilde{Z}}\tilde{X}\hat{\gamma}^u \\
&\qquad - (\Delta s^\top P_{\tilde{Z}}\Delta s - 2\Delta s^\top P_{\tilde{Z}}\tilde{X}\gamma^u_{2SLS} + \gamma^u_{2SLS}{}^\top\tilde{X}^\top P_{\tilde{Z}}\tilde{X}\gamma^u_{2SLS})| \\
&= \frac{1}{T}|2\Delta s^\top P_{\tilde{Z}}\tilde{X}(\gamma^u_{2SLS} - \hat{\gamma}^u) + \hat{\gamma}^{u\top}\tilde{X}^\top P_{\tilde{Z}}\tilde{X}\hat{\gamma}^u - \gamma^u_{2SLS}{}^\top\tilde{X}^\top P_{\tilde{Z}}\tilde{X}\gamma^u_{2SLS}| \\
&= \frac{1}{T}|2\Delta s^\top P_{\tilde{Z}}\tilde{X}(\gamma^u_{2SLS} - \hat{\gamma}^u) \\
&+ (\hat{\gamma}^u - \gamma^u_{2SLS} + \gamma^u_{2SLS})^\top\tilde{X}^\top P_{\tilde{Z}}\tilde{X}(\hat{\gamma}^u - \gamma^u_{2SLS} + \gamma^u_{2SLS}) - \gamma^u_{2SLS}{}^\top\tilde{X}^\top P_{\tilde{Z}}\tilde{X}\gamma^u_{2SLS}| \\
&= \frac{1}{T}|2\Delta s^\top P_{\tilde{Z}}\tilde{X}(\gamma^u_{2SLS} - \hat{\gamma}^u) + (\hat{\gamma}^u - \gamma^u_{2SLS})^\top\tilde{X}^\top P_{\tilde{Z}}\tilde{X}(\hat{\gamma}^u - \gamma^u_{2SLS}) \\
&\qquad + 2\gamma^u_{2SLS}{}^\top\tilde{X}^\top P_{\tilde{Z}}\tilde{X}(\hat{\gamma}^u - \gamma^u_{2SLS})| \\
&\geq C_3\|\hat{\gamma}^u - \gamma^u_{2SLS}\|_2,
\end{aligned}
$$

where $C_3 > 0$ is a constant. The last inequality is from the CS inequality and the boundness of the eigenvalues. Therefore, we obtain

$$\mathbb{P}(\|\hat{\gamma}^u - \gamma^u_{2SLS}\|_2 > \epsilon) < C_1T(e^{-C_2N\epsilon^2/J\log^2 J} + e^{-C_2N/J})$$

with positive constants C_1 and C_2. □

Lemma C.3. *Given any non-separating alternatives j and j' define $\hat{\gamma}^u = (\hat{\beta}, \hat{\alpha}, \hat{u})$ as estimates of the 2SLS procedure given in Algorithm 1. If $J = o(N)$, $\delta \leq \frac{c_0}{2J}$ and Assumption 5 holds, then with $\gamma^u := (\beta, \alpha, u)$, for any $0 < \epsilon < 2$*

$$\mathbb{P}(\|\hat{\gamma}^u - \gamma^u\|_2 > \epsilon) < C_1T(e^{-C_2N\epsilon^2/J\log^2 J} + e^{-C_2N/J}) + \tilde{C}_1e^{-\tilde{C}_2T\epsilon^2}$$

with positive constants $\tilde{C}_1$, $\tilde{C}_2$, C_1, and C_2.

Proof. This comes directly from the union bound. □

This Lemma contains three errors: first, in the quadratic approximation of the logarithm of empirical shares to actual shares when empirical shares are bounded; second, when the logarithmic share approaches zero; and third, in the errors arising from the Oracle 2SLS estimator.

D Selection of pure alternatives

In this section, we discuss the statistical guarantees of choosing non-separating alternatives. Our proof is new in that we check the non-asymptotic properties of these nonparametric tests.

The following lemmas analyze the convergence of observed tests under the null and alternative. To simplify, we drop jj'. Consider the leave-one-out kernel estimator:

$$\hat{\Phi} = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \hat{\xi}_t \hat{\xi}_\tau K_{h,t\tau}, \tag{11}$$

where $\hat{\xi}_t = \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \delta\}(\Delta\hat{s}_t - x_t^\top \hat{\beta} - \hat{\alpha}^{jj'} p_t - \hat{u})$, and $K_{h,t\tau} = h_1^{-1} h_2^{-1} k\left(\frac{z_{jt} - z_{j\tau}}{h_1}\right) k\left(\frac{z_{j't} - z_{j'\tau}}{h_2}\right)$.

We compare this to the Oracle test:

$$\hat{\Phi}^* = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \hat{\xi}_t^* \hat{\xi}_\tau^* K_{h,t\tau}, \tag{12}$$

where $\hat{\xi}_t^* = \Delta s_t - x_t^\top \hat{\beta} - \hat{\alpha} p_t - \hat{u}$.

The distance between

$$\mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \delta\}\mathbb{1}\{\hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} > \delta\}(\Delta\hat{s}_t - x_t^\top \hat{\beta} - \hat{\alpha} p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top \hat{\beta} - \hat{\alpha} p_\tau - \hat{u}) K_{h,t\tau}$$

and

$$(\Delta s_t - x_t^\top \hat{\beta} - \hat{\alpha} p_t - \hat{u})(\Delta s_\tau - x_\tau^\top \hat{\beta} - \hat{\alpha} p_\tau - \hat{u}) K_{h,t\tau}$$

has two components:

$$\begin{aligned} \tilde{H}_1 = & \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} \wedge \hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} > \delta\} \Big[(\Delta\hat{s}_t - x_t^\top \hat{\beta} - \hat{\alpha} p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top \hat{\beta} - \hat{\alpha} p_\tau - \hat{u}) \\ & - (\Delta s_t - x_t^\top \hat{\beta} - \hat{\alpha} p_t - \hat{u})(\Delta s_\tau - x_\tau^\top \hat{\beta} - \hat{\alpha} p_\tau - \hat{u})\Big] K_{h,t\tau}, \end{aligned}$$

and

$$\tilde{H}_2 = \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} \wedge \hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} < \delta\}(\Delta s_t - x_t^\top \hat{\beta} - \hat{\alpha} p_t - \hat{u})(\Delta s_\tau - x_\tau^\top \hat{\beta} - \hat{\alpha} p_\tau - \hat{u}) K_{h,t\tau}.$$

The following two lemmas help investigate this first part $\tilde{H}_1$.

Lemma D.1. *Given alternatives* $(j, j') \in [J]$ *and markets* $(t, \tau) \in [T]$ *with bounded utilities such* *that* $\frac{h_0}{J} < s_{il} < \frac{h_1}{J}$ *with any* $i \in \{j, j'\}$ *and* $l \in \{t, \tau\}$, *for any* $0 < \epsilon < 2$

$$\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta s_\tau - x_\tau\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})| > \epsilon] \leq C_1(e^{-C_2N\epsilon^2/J\log^2 J} + e^{-C_2N/J})$$

where C_1 *and* C_2 *are positive constants.*

Proof. We have

$$|(\Delta\hat{s}_t - \Delta s_t)(\Delta s_\tau - x_\tau\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})| \leq \tilde{C}_1 \log J|\Delta\hat{s}_t - \Delta s_t|$$

since x_τ, p_τ, $\hat{\beta}$, $\hat{\alpha}$, and $\hat{u}$ are bounded. $|\Delta\hat{s}_\tau| = |\log s_{j\tau} - \log s_{j'\tau}| \leq |\log s_{j\tau}| + |\log s_{j'\tau}| \leq \frac{2h_1}{J}$ where h_1 represents for the upper bound of the utility. There exists positive $\tilde{C}_1$ such that $|\Delta s_\tau - x_\tau\hat{\beta} - \hat{\alpha}p_\tau - \hat{u}| \leq \tilde{C}_1 \log J$.

Moreover, since

$$|\Delta\hat{s}_t - \Delta s_t| \leq |\log\hat{s}_{jt} - \log s_{jt}| + |\log\hat{s}_{j't} - \log s_{j't}|,$$

we can base our concentration inequality on

$$\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta s_\tau - x_\tau\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})| > \epsilon] \lesssim \mathbb{P}[\tilde{C}_1 \log J|\log\hat{s}_{jt} - \log s_{jt}| > \frac{1}{2}\epsilon]$$

by the union bound. Then, we rely on the Lemma F.21

$$\mathbb{P}[\log\hat{s}_{jt} - \log s_{jt}| > \frac{\epsilon}{2\tilde{C}_1 \log J}] \leq \bar{C}_1(e^{-\bar{C}_2N\epsilon^2/J\log^2 J} + e^{-\bar{C}_2N/J}).$$

Therefore,

$$\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta s_\tau - x_\tau\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})| > \epsilon] \leq C_1(e^{-C_2N\epsilon^2/J\log^2 J} + e^{-C_2N/J})$$

□

Lemma D.2. *Given alternatives* $(j, j') \in [J]$ *and markets* $(t, \tau) \in [T]$ *with bounded utilities such that* $\frac{h_0}{J} < s_{il} < \frac{h_1}{J}$ *with any* $i \in \{j, j'\}$ *and* $l \in \{t, \tau\}$, *for any* $0 < \epsilon < 2$

$$\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)| > \epsilon] \leq C_1(e^{-C_2N\epsilon^2/J\log^2 J} + e^{-C_2N/J}).$$

Proof. Decompose

$$\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)| > \epsilon] = \underbrace{\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)| > \epsilon \mid \hat{s}_{j\tau} \wedge \hat{s}_{j'} > \frac{c}{2J}]\mathbb{P}(\hat{s}_{j\tau} \wedge \hat{s}_{j'} > \frac{c}{2J})}_{H_1}$$
$$\underbrace{\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)| > \epsilon \mid \hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} < \frac{c}{2J}]\mathbb{P}(\hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} < \frac{c}{2J})}_{H_2}.$$

For H_1,

$$\begin{aligned}
H_1 &\leq \mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)| > \epsilon \mid \hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} > \frac{c_0}{2J}] \\
&\leq \mathbb{P}[\tilde{C}_1 \log J|\Delta\hat{s}_t - \Delta s_t| > \epsilon \mid \hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} > \frac{c_0}{2J}] \\
&\leq \frac{\mathbb{P}[\tilde{C}_1 \log J|\Delta\hat{s}_t - \Delta s_t| > \epsilon]}{\mathbb{P}(\hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} > \frac{c_0}{2J})} \lesssim \mathbb{P}[\tilde{C}_1 \log J|\Delta\hat{s}_t - \Delta s_t| > \epsilon]
\end{aligned}$$

We therefore have $H_1 \leq \bar{C}_1(e^{-\bar{C}_2 N\epsilon^2/J\log^2 J} + e^{-\tilde{C}_2 N/J})$.

For H_2,

$$\mathbb{P}(\hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} < \frac{c_0}{2J}) \leq \bar{C}_1 e^{-\bar{C}_2 N/J}$$

by Lemma F.20 and the union bound. Therefore we have

$$\mathbb{P}[|(\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)| > \epsilon] \leq C_1(e^{-C_2 N\epsilon^2/J\log^2 J} + e^{-C_2 N/J}).$$

for some positive constant C_1 and C_2. □

Lemma D.3. *Given alternatives $(j, j') \in [J]$ with bounded utilities such that $\frac{h_0}{J} < s_{it} < \frac{h_1}{J}$ with any $i \in \{j, j'\}$ and $t \in [T]$, for any $0 < \epsilon < 2$*

$$\mathbb{P}[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T} \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \delta\}\mathbb{1}\{\hat{s}_{j\tau} \wedge \hat{s}_{j'\tau} > \delta\}|(\Delta\hat{s}_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})K_{h,t\tau}$$
$$- (\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta s_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})K_{h,t\tau}| > \epsilon]$$
$$\leq C_1 T^2(e^{-C_2 N h_1^2 h_2^2\epsilon^2/J\log^2 J} + e^{-C_2 N/J})$$

where C_1 and C_2 are positive constants.

Proof.

$$
\begin{aligned}
&\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\left(\Delta\hat{s}_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u}\right)\left(\Delta\hat{s}_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u}\right)K_{h,t\tau}.\\
&= \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\left(\Delta\hat{s}_t - \Delta s_t + \Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u}\right)\left(\Delta\hat{s}_\tau - \Delta s_\tau + \Delta s_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u}\right)K_{h,t\tau}\\
&= \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}[(\Delta\hat{s}_t - \Delta s_t + \Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta\hat{s}_\tau - \Delta s_\tau + \Delta s_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})\\
&\qquad + (\Delta\hat{s}_\tau - \Delta s_\tau)(\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u}) + (\Delta\hat{s}_t - \Delta s_t)(\Delta s_\tau - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})\\
&\qquad + (\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)]K_{h,t\tau}\\
&= \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}[(\Delta\hat{s}_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})\\
&\qquad + (\Delta\hat{s}_\tau - \Delta s_\tau)(\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u}) + (\Delta\hat{s}_t - \Delta s_t)(\Delta s_\tau - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})\\
&\qquad + (\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)]K_{h,t\tau}.
\end{aligned}
$$

where $K_{h,t\tau} = h_1^{-1}h_2^{-1}k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2})$. Therefore we have

$$
\begin{aligned}
&\mathbb{P}[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't} > \delta\}\mathbb{1}\{\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau} > \delta\}|(\Delta\hat{s}_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}^{jj'}p_\tau - \hat{u})K_{h,t\tau}\\
&\qquad\qquad - (\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta s_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})K_{h,t\tau}| > \epsilon]\\
&\leq \mathbb{P}[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}|[(\Delta\hat{s}_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})\\
&\qquad\qquad - (\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta s_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})]K_{h,t\tau}| > \epsilon]\\
&\lesssim T^2\mathbb{P}[|(\Delta\hat{s}_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})\\
&\qquad\qquad - (\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta s_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})]K_{h,t\tau}| > \epsilon]\\
&\leq T^2\mathbb{P}[|(\Delta\hat{s}_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta\hat{s}_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})\\
&\qquad\qquad - (\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})(\Delta s_\tau - x_\tau^\top\hat{\beta} - \hat{\alpha}p_\tau - \hat{u})]| > h_1h_2\epsilon]\\
&\leq T^2\mathbb{P}[|(\Delta\hat{s}_\tau - \Delta s_\tau)(\Delta s_t - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u}) + (\Delta\hat{s}_t - \Delta s_t)(\Delta s_\tau - x_t^\top\hat{\beta} - \hat{\alpha}p_t - \hat{u})\\
&\qquad + (\Delta\hat{s}_t - \Delta s_t)(\Delta\hat{s}_\tau - \Delta s_\tau)| > h_1h_2\epsilon]\\
&\leq C_1T^2(e^{-C_2Nh_1^2h_2^2\epsilon^2/J\log^2 J} + e^{-C_2N/J})
\end{aligned}
$$

where the first approximated inequality is from the union bound. The first inequality is from the kernel function, which is bounded. The second inequality is from the argument above, and the final inequality is from the Lemmas. $\square$

Lemma D.4. *Given alternatives $(j, j') \in [J]$ and markets $(t, \tau) \in [T]$ with bounded utilities such that $\frac{h_0}{J} < s_{il} < \frac{h_1}{J}$ with any $i \in \{j, j'\}$ and $l \in \{t, \tau\}$, assume $0 < \delta < \frac{h_0}{2J}$, and $J = o(N)$. For any $\epsilon > 0$*

$$\mathbb{P}[\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}<\delta\}(\Delta s_t - x_t^\top\hat{\beta}-\hat{\alpha}p_t-\hat{u})(\Delta s_\tau - x_\tau^\top\hat{\beta}-\hat{\alpha}p_\tau-\hat{u})K_{h,t\tau}| > \epsilon]$$
$$\leq C_1T^2e^{-C_2N/J}$$

Proof. We condition on the event $\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\frac{h_0}{2J}$ and $\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}>\frac{h_0}{2J}$.

$$\begin{aligned}
&\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\delta\}(\Delta s_t - x_t^\top\hat{\beta}-\hat{\alpha}p_t-\hat{u})(\Delta s_\tau - x_\tau^\top\hat{\beta}-\hat{\alpha}p_\tau-\hat{u})K_{h,t\tau}|>\epsilon)\\
&=\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\delta\}(\Delta s_t - x_t^\top\hat{\beta}-\hat{\alpha}p_t-\hat{u})\\
&\qquad\times(\Delta s_\tau - x_\tau^\top\hat{\beta}-\hat{\alpha}p_\tau-\hat{u})K_{h,t\tau}|>\epsilon\mid\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\frac{h_0}{2J})\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\frac{h_0}{2J})\\
&+\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\delta\}(\Delta s_t - x_t^\top\hat{\beta}-\hat{\alpha}p_t-\hat{u})\\
&\qquad\times(\Delta s_\tau - x_\tau^\top\hat{\beta}-\hat{\alpha}p_\tau-\hat{u})K_{h,t\tau}|>\epsilon\mid\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}>\frac{h_0}{2J})\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}>\frac{h_0}{2J})\\
&=\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\delta\}(\Delta s_t - x_t^\top\hat{\beta}-\hat{\alpha}p_t-\hat{u})\\
&\qquad\times(\Delta s_\tau - x_\tau^\top\hat{\beta}-\hat{\alpha}p_\tau-\hat{u})K_{h,t\tau}|>\epsilon\mid\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\frac{h_0}{2J})\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\frac{h_0}{2J})\\
&\leq\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}\wedge\hat{s}_{j\tau}\wedge\hat{s}_{j'\tau}\leq\frac{h_0}{2J})\lesssim e^{-h_0^2N/8c_1J}
\end{aligned}$$

where the second equality is from $\delta\leq\frac{h_0}{2J}$ and the last inequality is by Lemma F.20. $\square$

Lemma D.5. *Given alternatives $(j, j') \in [J]$ and markets $(t, \tau) \in [T]$ with bounded utilities such that $\frac{h_0}{J} < s_{il} < \frac{h_1}{J}$ with any $i \in \{j, j'\}$ and $l \in \{t, \tau\}$, for any $\epsilon > 0$*

$$\mathbb{P}[|\hat{\Phi}-\hat{\Phi}^*|>\epsilon]\leq C_1T^2(e^{-C_2Nh_1^2h_2^2\epsilon^2/J\log^2 J}+e^{-C_2N/J})$$

Proof. The proof is directly from the union bound with Lemma D.3 and Lemma D.4. $\square$

Then, we bound $\hat{\Phi}^*$ for both the null and alternative hypotheses:

$$\hat{\Phi}^* = \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\hat{\xi}_t^*\hat{\xi}_\tau^*h_1^{-1}h_2^{-1}k\left(\frac{z_{jt}-z_{j\tau}}{h_1}\right)k\left(\frac{z_{j't}-z_{j'\tau}}{h_2}\right),$$

where $\hat{\xi}_t^* = \Delta s_t - x_t^\top\hat{\beta}-\hat{\alpha}p_t-\hat{u}$.

Under the null hypothesis:

$$\hat{\Phi}^* = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \xi_t \xi_t K_{h,t\tau} - \frac{2}{T(T-1)} (\hat{\gamma}^u - \gamma^u)^\top \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \xi_t \tilde{x}_\tau K_{h,t\tau}$$
$$+ \frac{1}{T(T-1)} (\hat{\gamma}^u - \gamma^u)^\top \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \tilde{x}_t \tilde{x}_\tau K_{h,t\tau} (\hat{\gamma}^u - \gamma^u),$$

where $\Delta s_t = \tilde{x}^\top \gamma^u + \xi_t$.

We will show that the test converges to zero exponentially under the null, with the tail bound satisfying $e^{-r(T)/\epsilon}$, where $r(T)/\epsilon$ is the rate of convergence. The first term is a U-statistic with mean zero, and the second and third terms are bounded by $(\hat{\gamma}^u - \gamma^u)$ whose rates have been discussed in the lemmas above.

On the other hand, under the alternative hypothesis:

$$\hat{\Phi}^* = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \left[\tilde{\xi}_t + g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})\right]$$
$$\times \left[\tilde{\xi}_\tau + g(z_{j\tau}, z_{j'\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{j\tau}, z_{j'\tau})\right] K_{h,t\tau}$$
$$- \frac{2}{T(T-1)} (\hat{\gamma}^u - \bar{\gamma}^u)^\top \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \left[\tilde{\xi}_t + g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})\right] \tilde{x}_\tau K_{h,t\tau}$$
$$+ \frac{1}{T(T-1)} (\hat{\gamma}^u - \bar{\gamma}^u)^\top \sum_{t=1}^{T} \sum_{\tau=1,\tau\neq t}^{T} \mathbb{E}(\tilde{x}_t \mid z_{jt}, z_{j't}) \mathbb{E}(\tilde{x}_\tau^\top \mid z_{j\tau}, z_{j'\tau}) K_{h,t\tau} (\hat{\gamma}^u - \bar{\gamma}^u),$$

where $\Delta s_t = g(\tilde{x}_t) + \bar{\xi}_t$ and $g(z_{jt}, z_{j't}) = \mathbb{E}(\Delta s_t \mid z_{jt}, z_{j't})$. We aim to show that this converges to:

$$\mathbb{E}\left\{\left[\mathbb{E}(\Delta s_t \mid z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})\right]^2 f(z_{jt}, z_{j't})\right\} \geq c_s,$$

where c_s is a positive constant. The parameter $\bar{\gamma}^u$ ensures that the 2SLS still converges, and a tuning parameter between zero and a constant can distinguish separating from non-separating alternatives.

The following lemmas focus on the tail bounds for the null hypothesis, starting with the Oracle case and proceeding to the formal test.

Lemma D.6. *Under the null hypothesis—that is, for non-separating alternatives j and j' with bounded utilities such that $\frac{h_0}{J} < s_{it} < \frac{h_1}{J}$ with any $i \in \{j, j'\}$ and $t \in [T]$— define $\hat{\gamma}^u = (\hat{\beta}, \hat{\alpha}, \hat{u})$ as* *estimates of the 2SLS procedure given in Algorithm 1. If $J = o(N)$ and $\delta \leq \frac{c_0}{2J}$ for any* $0 < \epsilon < 1$

$$\mathbb{P}(|\hat{\Phi}^*| > \epsilon) \leq C_1 T(e^{-C_2 h_1 h_2 N\epsilon^2/J\log^2 J} + e^{-C_2 N/J}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1 h_2 T\epsilon^2}.$$

Proof. We use the equality:

$$\hat{\Phi}^* = \underbrace{\frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\xi_t\xi_\tau K_{h,t\tau}}_{H_1} - \underbrace{\frac{2}{T(T-1)}(\hat{\gamma}^u - \gamma^u)^\top \sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\xi_t\tilde{x}_\tau K_{h,t\tau}}_{H_2} + \underbrace{\frac{1}{T(T-1)}(\hat{\gamma}^u - \gamma^u)^\top \sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\tilde{x}_t\tilde{x}_\tau^\top K_{h,t\tau}(\hat{\gamma}^u - \gamma^u)}_{H_3}$$

For the H_1, noticing that it is a U-statistic, we can directly use Lemma F.12, which is an application of Heoffding's decomposition and characteristic function. We also have

$$\begin{aligned}
\mathbb{E}[\xi_t\xi_\tau h_1^{-1}h_2^{-1}k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2})] &= h_1^{-1}h_2^{-1}\mathbb{E}[\xi_t\xi_\tau k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2})] \\
&= h_1^{-1}h_2^{-1}\mathbb{E}\{\mathbb{E}[\xi_t\xi_\tau k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2}) \mid z_t, z_\tau]\} \\
&= h_1^{-1}h_2^{-1}\mathbb{E}\{k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2})\mathbb{E}[\xi_t\xi_\tau \mid z_t, z_\tau]\} \\
&= 0,
\end{aligned}$$

where the second equality uses the law of iterated expectation. The last equality is from the exclusion restriction. Therefore,

$$\begin{aligned}
\mathbb{P}(|H_1| > \epsilon) &\lesssim e^{-T\epsilon^2/(4k\sigma^2+\frac{2}{3}kM\epsilon)} \leq e^{-T\epsilon^2/(8\sigma^2+\frac{4}{3}h_1^{-1}h_2^{-1}\epsilon)} \leq e^{-T\epsilon^2/(8\kappa_1 h_1^{-1}h_2^{-1}+\frac{4}{3}h_1^{-1}h_2^{-1}\epsilon)} \\
&\leq e^{-T\epsilon^2/(8\kappa_1 h_1^{-1}h_2^{-1}+\frac{8}{3}h_1^{-1}h_2^{-1})} \leq e^{-\tilde{C}_2 T h_1 h_2\epsilon^2}
\end{aligned}$$

where the first inequality is from $k=2$. Also,

$$
\begin{aligned}
\sigma^2 &= \mathbb{E}[\xi_t\xi_\tau h_1^{-1}h_2^{-1}k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2})]^2 \\
&= \mathbb{E}\{\mathbb{E}[\xi_t\xi_\tau h_1^{-1}h_2^{-1}k(\frac{z_{jt}-z_{j\tau}}{h_1})k(\frac{z_{j't}-z_{j'\tau}}{h_2}) \mid z_{jt}, z_{j't}, z_{j\tau}, z_{j\tau}]^2\} \\
&= \mathbb{E}\{h_1^{-2}h_2^{-2}k^2(\frac{z_{jt}-z_{j\tau}}{h_1})k^2(\frac{z_{j't}-z_{j'\tau}}{h_2})\mathbb{E}[\xi_t\xi_\tau \mid z_{jt}, z_{j't}, z_{j\tau}, z_{j\tau}]^2\} \\
&= h_1^{-2}h_2^{-2}\iiiint f(z_{jt}, z_{j't})f(z_{j\tau}, z_{j'\tau})\sigma^2(z_{jt}, z_{j't})\sigma^2(z_{j\tau}, z_{j'\tau}) \\
&\quad \times k^2(\frac{z_{jt}-z_{j\tau}}{h_1})k^2(\frac{z_{j't}-z_{j'\tau}}{h_2})dz_{jt}dz_{j't}dz_{j\tau}dz_{j'\tau} \\
&= h_1^{-1}h_2^{-1}\iiiint f(z_{jt}, z_{j't})f(z_{jt}+h_1u, z_{j't}+h_2v)\sigma^2(z_{jt}, z_{j't})\sigma^2(z_{jt}+h_1u, z_{j't}+h_2v) \\
&\quad \times k^2(u)\,k^2(v)\,dz_{jt}dz_{j't}dudv \\
&= h_1^{-1}h_2^{-1}\iiiint f(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't})\Bigg[f(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't}) \\
&\quad + \left(\frac{\partial f}{\partial z_{jt}}(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial\sigma^2}{\partial z_{jt}}(z_{jt}, z_{j't})\right)h_1u \\
&\quad + \left(\frac{\partial f}{\partial z_{j't}}(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial\sigma^2}{\partial z_{j't}}(z_{jt}, z_{j't})\right)h_2v \\
&\quad + \frac{1}{2}\left(\frac{\partial^2 f}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2\sigma^2}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\right)h_1^2u^2 \\
&\quad + \frac{1}{2}\left(\frac{\partial^2 f}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2\sigma^2}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\right)h_2^2v^2 \\
&\quad + \left(\frac{\partial^2 f}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2\sigma^2}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})\right)h_1h_2uv\Bigg] \\
&\quad \times k^2(u)k^2(v)\,dudvdz_{jt}dz_{j't}
\end{aligned}
$$

by change of variables that $z_{j\tau} = z_{jt} + h_1u$ and $z_{j'\tau} = z_{j\tau} + h_2v$. The sixth inequality is from the Taylor expansion. $\tilde{z}_{jt}$ lies between z_{jt} and $z_{j\tau}$, with $\tilde{z}_{j't}$ between $z_{j't}$ and $z_{j'\tau}$. This can be further expressed as:

$$
\begin{aligned}
\sigma^2 = h_1^{-1}h_2^{-1} & \iiiint f(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't}) \\
& \times \Bigg[f(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't}) + 0 + 0 + \frac{1}{2}\left(\frac{\partial^2 f}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2 \sigma^2}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't}) \right) h_1^2 u^2 \\
& + \frac{1}{2}\left(\frac{\partial^2 f}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2 \sigma^2}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't}) \right) h_2^2 v^2 \\
& + \left(\frac{\partial^2 f}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2 \sigma^2}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't}) \right) h_1 h_2 uv \Bigg] k^2(u)k^2(v)\, dz_{jt}dz_{j't}dudv \\
= {} & h_1^{-1}h_2^{-1} \iint f^2(z_{jt}, z_{j't})\sigma^4(z_{jt}, z_{j't})dz_{jt}dz_{j't} \iint k^2(u)k^2(v)dudv \\
& + h_1 h_2^{-1} \iiiint \frac{1}{2} f(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't}) \\
& \qquad \times \left(\frac{\partial^2 f}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2 \sigma^2}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't}) \right) u^2 k^2(u)k^2(v)dudvdz_{jt}dz_{j't} \\
& + h_1^{-1} h_2 \iiiint \frac{1}{2} f(z_{jt}, z_{j't})\sigma^2(z_{jt}, z_{j't}) \\
& \qquad \times \left(\frac{\partial^2 f}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2 \sigma^2}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't}) \right) k^2(u)v^2 k^2(v)dudvdz_{jt}dz_{j't} \\
& + \iiiint \left(\frac{\partial^2 f}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})\sigma^2(z_{jt}, z_{j't}) + f(z_{jt}, z_{j't})\frac{\partial^2 \sigma^2}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't}) \right) uk^2(u)vk^2(v)dudvdz_{jt}dz_{j't} \\
\lesssim {} & h_1^{-1}h_2^{-1},
\end{aligned}
$$

where we also use the CS inequality that $\int k^2(u)du \leq [\int k(u)du]^2 = 1$, $\int uk^2(u)du = 0$ by $k(u) = k(-u)$. We cannot take the integral of (u, v) out since the second derivatives depend on z and (u, v).

Then, we discuss H_2. First of all, we notice

$$
\begin{aligned}
\mathbb{E}[\xi_t x_\tau h_1^{-1}h_2^{-1} k(\frac{z_{jt} - z_{j\tau}}{h_1})k(\frac{z_{j't} - z_{j'\tau}}{h_2})] &= h_1^{-1}h_2^{-1}\mathbb{E}[\xi_t x_\tau k(\frac{z_{jt} - z_{j\tau}}{h_1})k(\frac{z_{j't} - z_{j'\tau}}{h_2})] \\
&= h_1^{-1}h_2^{-1}\mathbb{E}\{\mathbb{E}[\xi_t x_\tau k(\frac{z_{jt} - z_{j\tau}}{h_1})k(\frac{z_{j't} - z_{j'\tau}}{h_2}) \mid z_t, z_\tau, x_\tau]\} \\
&= h_1^{-1}h_2^{-1}\mathbb{E}\{x_\tau k(\frac{z_{jt} - z_{j\tau}}{h_1})k(\frac{z_{j't} - z_{j'\tau}}{h_2})\mathbb{E}[\xi_t \mid z_t, z_\tau, x_\tau]\} \\
&= 0.
\end{aligned}
$$

We can also get

$$\mathbb{P}(\|\frac{2}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\xi_t\tilde{x}_\tau K_{h,t\tau}\|_2>\epsilon)\leq C_{2,1}e^{-C_{2,2}Th_1h_2\epsilon^2}$$

by the same way as H_1. Therefore, H_2 has

$$\begin{aligned}
&\mathbb{P}[|(\hat{\gamma}^u-\gamma^u)^\top\frac{2}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\xi_t\tilde{x}_\tau K_{h,t\tau}|>\epsilon]\\
&\leq\mathbb{P}(\|\hat{\gamma}^u-\gamma^u\|_2\|\frac{2}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\xi_t\tilde{x}_\tau K_{h,t\tau}\|_2>\epsilon)\\
&\leq\mathbb{P}(\|\hat{\gamma}^u-\gamma^u\|_2>\sqrt{\epsilon})+\mathbb{P}(\|\frac{2}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\xi_t\tilde{x}_\tau K_{h,t\tau}\|_2>\sqrt{\epsilon})\\
&\leq\mathbb{P}(\|\hat{\gamma}^u-\gamma^u\|_2>\epsilon)+\mathbb{P}(\|\frac{2}{T(T-1)}\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\xi_t\tilde{x}_\tau K_{h,t\tau}\|_2>\epsilon)\\
&\leq C_1T(e^{-C_2N\epsilon^2/J\log^2J}+e^{-C_2N/J})+\tilde{C}_1e^{-\tilde{C}_2T\epsilon^2}+e^{-C_2N/J})+\tilde{C}_1e^{-\tilde{C}_2Th_1h_2\epsilon^2}\\
&\leq C_1T(e^{-C_2N\epsilon^2/J\log^2J}+e^{-C_2N/J})+\tilde{C}_1e^{-\tilde{C}_2Th_1h_2\epsilon^2}
\end{aligned}$$

where the first inequality is from the CS inequality. The second inequality is from the union bound and the third is from $\epsilon<1$. The forth inequality is directly based on Lemma C.3.

Finally, we focus on H_3:

$$|\frac{1}{T(T-1)}(\hat{\gamma}^u-\gamma^u)^\top\sum_{t=1}^{T}\sum_{\tau=1,\tau\neq t}^{T}\tilde{x}_t\tilde{x}_\tau^\top K_{h,t\tau}(\hat{\gamma}^u-\gamma^u)|\lesssim h_1^{-1}h_2^{-1}\|\hat{\gamma}^u-\gamma^u\|_2^2.$$

We get

$$\begin{aligned}
\mathbb{P}(|H_3|>\epsilon)&\leq\mathbb{P}(\kappa_2h_1^{-1}h_2^{-1}\|\hat{\gamma}^u-\gamma^u\|_2^2>\epsilon)=\mathbb{P}(\|\hat{\gamma}^u-\gamma^u\|_2>\sqrt{\frac{h_1h_2\epsilon}{\kappa_2}})\\
&\leq\mathbb{P}(\|\hat{\gamma}^u-\gamma^u\|_2>\sqrt{\frac{h_1h_2}{\kappa_2}}\epsilon)\leq C_1T(e^{-C_2h_1h_2N\epsilon^2/J\log^2J}+e^{-C_2N/J})+\tilde{C}_1e^{-\tilde{C}_2h_1h_2T\epsilon^2}
\end{aligned}$$

by $\epsilon<1$.

In conclusion, for any $0<\epsilon<1$ we obtain the bound:

$$\mathbb{P}(|\hat{\Phi}^*|>\epsilon)\leq C_1T(e^{-C_2h_1h_2N\epsilon^2/J\log^2J}+e^{-C_2N/J})+\tilde{C}_1e^{-\tilde{C}_2h_1h_2T\epsilon^2}.$$

□

The rate here is different from the previous lemma in $h_1 h_2$ because the previous lemma did not have a mean zero.

Lemma D.7. *Under the null hypothesis—that is, for non-separating alternatives j and $j^{'}$ with bounded utilities such that $\frac{h_0}{J} < s_{it} < \frac{h_1}{J}$ with any $i \in \{j, j^{'}\}$ and $t \in [T]$— define $\hat{\gamma}^u = (\hat{\beta}, \hat{\alpha}, \hat{u})$ as estimates of the 2SLS procedure given in Algorithm 1. If $J = o(N)$ and $\delta \leq \frac{c_0}{2J}$ for any* $0 < \epsilon < 1$

$$\mathbb{P}(|\hat{\Phi}| > \epsilon) \leq C_1 T^2 (e^{-C_2 h_1^2 h_2^2 N \epsilon^2 / J \log^2 J} + e^{-C_2 N/J}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1 h_2 T \epsilon^2}.$$

Proof. This comes from the union bound. □

This lemma focuses on the tail bound for the alternative hypothesis. While we could still apply Bernstein's inequality, it is unnecessary here, as the rate for the alternative does not need to be strict to achieve rate consistency. However, we need to expand it more to check the

Lemma D.8. *Under the alternative hypothesis—that is, for separating alternatives j and $j^{'}$ with bounded utilities such that $\frac{h_0}{J} < s_{it} < \frac{h_1}{J}$ with any $i \in \{j, j^{'}\}$ and $t \in [T]$— define $\hat{\gamma}^u = (\hat{\beta}, \hat{\alpha}, \hat{u})$ as estimates of the 2SLS procedure given in Algorithm 1. If $J = o(N)$ and $\delta \leq \frac{c_0}{2J}$ for any* $0 < \epsilon < 1$ *then there exists $O(h_1^2)$ and $O(h_2^2)$ such that*

$$\begin{aligned}&\mathbb{P}(|\hat{\Phi}^* - \mathbb{E}\Big\{\big[g(z_{jt}, z_{j^{'}t}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j^{'}t})\big]^2 f(z_{jt}, z_{j^{'}t})\Big\} - O(h_1^2) - O(h_2^2)| > \epsilon) \\ &\leq \tilde{C}_1 e^{-\tilde{C}_2 h_1^2 h_2^2 T \epsilon^2}.\end{aligned}$$

Proof.

$$\begin{aligned}\hat{\Phi}^* = &\frac{1}{T(T-1)} \sum_{t=1}^T \sum_{\tau=1, \tau \neq t}^T [\tilde{\xi}_t + g(z_{jt}, z_{j^{'}t}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u | z_{jt}, z_{j^{'}t})][\tilde{\xi}_\tau + g(z_{j\tau}, z_{j^{'}\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u | z_{j\tau}, z_{j^{'}\tau})] K_{h,t\tau} \\ &- \frac{2}{T(T-1)} (\hat{\gamma}^u - \bar{\gamma}^u)^\top \sum_{t=1}^T \sum_{\tau=1, \tau \neq t}^T [\tilde{\xi}_t + g(z_{jt}, z_{j^{'}t}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u | z_{jt}, z_{j^{'}t})] \tilde{x}_\tau K_{h,t\tau} \\ &+ \frac{1}{T(T-1)} (\hat{\gamma}^u - \bar{\gamma}^u)^\top \sum_{t=1}^T \sum_{\tau=1, \tau \neq t}^T \mathbb{E}(\tilde{x}_t \mid z_{jt}, z_{j^{'}t}) \mathbb{E}(\tilde{x}_\tau^\top \mid z_{j\tau}, z_{j^{'}\tau}) K_{h,t\tau} (\hat{\gamma}^u - \bar{\gamma}^u)\end{aligned}$$

We focus on the first part

$$U_T = \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{\tau=1, \tau \neq t}^T [\tilde{\xi}_t + g(z_{jt}, z_{j^{'}t}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u | z_{jt}, z_{j^{'}t})][\tilde{\xi}_\tau + g(z_{j\tau}, z_{j^{'}\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u | z_{j\tau}, z_{j^{'}\tau})] K_{h,t\tau}.$$

It is a U-statistic, therefore, by the Lemma F.11 of Hoeffding:

$$P(|U_T - \mathbb{E}(U_T)| > \epsilon) < C_{1,1} e^{-C_{1,2} T h_1^2 h_2^2 \epsilon^2}.$$

We calculate the expectation

$$\begin{aligned}
\mathbb{E}(U_n) &= \mathbb{E}[\tilde{\xi}_t + g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u | z_{jt}, z_{j't})][\tilde{\xi}_\tau + g(z_{j\tau}, z_{j'\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u | z_{j\tau}, z_{j'\tau})] K_{h,t\tau} \\
&= \mathbb{E}\{[g(z_{j\tau}, z_{j'\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u | z_{j\tau}, z_{j'\tau})][g(z_{j\tau}, z_{j'\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u | z_{j\tau}, z_{j'\tau})] K_{h,t\tau}\}
\end{aligned}$$

by the law of iterated expectation and the exclusion restriction. Specifically, we want to check

$$\begin{aligned}
& h_1^{-1} h_2^{-1} \mathbb{E}\bigg\{ [g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})] \\
& \quad \times [g(z_{j\tau}, z_{j'\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{j\tau}, z_{j'\tau})] k\left(\frac{z_{jt} - z_{j\tau}}{h_1}\right) k\left(\frac{z_{j't} - z_{j'\tau}}{h_2}\right) \bigg\} \\
&= h_1^{-1} h_2^{-1} \iiiint f(z_{jt}, z_{j't}) f(z_{j\tau}, z_{j'\tau}) [g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})] \\
& \quad \times [g(z_{j\tau}, z_{j'\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{j\tau}, z_{j'\tau})] k\left(\frac{z_{jt} - z_{j\tau}}{h_1}\right) k\left(\frac{z_{j't} - z_{j'\tau}}{h_2}\right) dz_{jt} dz_{j't} dz_{j\tau} dz_{j'\tau} \\
&= \iiiint f(z_{jt}, z_{j't}) f(z_{jt} + h_1 u, z_{j't} + h_2 v) [g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})] \\
& \quad \times [g(z_{jt} + h_1 u, z_{j't} + h_2 v) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt} + h_1 u, z_{j't} + h_2 v)] k(-u) k(-v) du dv dz_{jt} dz_{j't} \\
&= \iiiint f(z_{jt}, z_{j't}) f(z_{jt} + h_1 u, z_{j't} + h_2 v) [g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})] \\
& \quad \times [g(z_{jt} + h_1 u, z_{j't} + h_2 v) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt} + h_1 u, z_{j't} + h_2 v)] k(u) k(v) du dv dz_{jt} dz_{j't},
\end{aligned}$$

where the first equality is by definition. The second equality is from the change of the variable that $z_{j\tau} = z_{jt} + h_1 u$ and $z_{j'\tau} = z_{j't} + h_2 v$ (h_1 and h_2 disappear because of it). In order to show that it is close to $\mathbb{E}\{[\mathbb{E}(\Delta s_t \mid z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u | z_{jt}, z_{j't})]^2 f(z_{jt}, z_{j't})\}$, we use second-order Taylor expansion to transform the model around z_{jt} and $z_{j't}$. We further simplify this equation to

$$
\begin{aligned}
= \iiiint & f(z_{jt}, z_{j't})\Big[f(z_{jt}, z_{j't}) + \frac{\partial f}{\partial z_{jt}}(z_{jt}, z_{j't})h_1 u + \frac{\partial f}{\partial z_{j't}}(z_{jt}, z_{j't})h_2 v \\
& + \frac{1}{2}\frac{\partial^2 f}{\partial z_{jt}^2}(z_{jt}, z_{j't})h_1^2 u^2 + \frac{\partial^2 f}{\partial z_{jt}\partial z_{j't}}(z_{jt}, z_{j't})h_1 h_2 uv + \frac{1}{2}\frac{\partial^2 f}{\partial z_{j't}^2}(z_{jt}, z_{j't})h_2^2 v^2\Big] \\
& \times \Big[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})\Big] \\
& \times \Big[g(z_{jt}, z_{j't}) + \frac{\partial g}{\partial z_{jt}}(z_{jt}, z_{j't})h_1 u + \frac{\partial g}{\partial z_{j't}}(z_{jt}, z_{j't})h_2 v \\
& + \frac{1}{2}\frac{\partial^2 g}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1^2 u^2 + \frac{\partial^2 g}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1 h_2 uv + \frac{1}{2}\frac{\partial^2 g}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_2^2 v^2 \\
& - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't}) - \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{jt}}h_1 u - \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{j't}}h_2 v \\
& - \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}^2}h_1^2 u^2 - \frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}\partial z_{j't}}h_1 h_2 uv \\
& - \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{j't}^2}h_2^2 v^2\Big]k(v)k(u)\, dz_{jt}\, dz_{j't}\, du\, dv.
\end{aligned}
$$

Then, we want to reorder the term for further analysis. Before that, we define

$$
\begin{aligned}
f &:= f(z_{jt}, z_{j't}) \\
g &:= g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't}).
\end{aligned}
$$

Also define

$$
\begin{aligned}
f^+ := & \frac{\partial f}{\partial z_{jt}}(z_{jt}, z_{j't})h_1 u + \frac{\partial f}{\partial z_{j't}}(z_{jt}, z_{j't})h_2 v \\
& + \frac{1}{2}\frac{\partial^2 f}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1^2 u^2 + \frac{\partial^2 f}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1 h_2 uv + \frac{1}{2}\frac{\partial^2 f}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_2^2 v^2,
\end{aligned}
$$

and

$$
\begin{aligned}
g^+ :=& \frac{\partial g}{\partial z_{jt}}(z_{jt}, z_{j't})h_1 u + \frac{\partial g}{\partial z_{j't}}(z_{jt}, z_{j't})h_2 v \\
&+ \frac{1}{2}\frac{\partial^2 g}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1^2 u^2 + \frac{\partial^2 g}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1 h_2 uv + \frac{1}{2}\frac{\partial^2 g}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_2^2 v^2 \\
&- \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{jt}}h_1 u - \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{j't}}h_2 v \\
&- \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}^2}h_1^2 u^2 - \frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}\partial z_{j't}}h_1 h_2 uv - \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{j't}^2}h_2^2 v^2.
\end{aligned}
$$

With these notations, the equation inside the integral that we are interested in is

$$
\begin{aligned}
fg(f + f^+)(g + g^+) &= (f^2 g + f f^+ g)(g + g^+) \\
&= f^2 g^2 + f^2 g g^+ + f f^+ g^2 + f f^+ g g^+,
\end{aligned}
$$

on which we want to show that $f^2 g g^+ + f f^+ g^2 + f f^+ g g^+$ is small, because

$$
\begin{aligned}
\iiiint f^2 g^2 k(u)k(v) du dv dz_{jt} dz_{j't} &= \iint f^2 g^2 dz_{jt} dz_{j't} \\
&= \mathbb{E}\Big\{\big[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})\big]^2 f(z_{jt}, z_{j't})\Big\}
\end{aligned}
$$

where the first equality is from $\iint k(u)k(v) du dv = 1$.

Then, for the second part,

$$
\begin{aligned}
&\iiiint f^2 g g^+ du dv dz_{jt} dz_{j't} \\
&= \iiiint f^2(z_{jt}, z_{j't})[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})] \\
&\quad \times [\frac{\partial g}{\partial z_{jt}}(z_{jt}, z_{j't})h_1 u + \frac{\partial g}{\partial z_{j't}}(z_{jt}, z_{j't})h_2 v \\
&\quad + \frac{1}{2}\frac{\partial^2 g}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1^2 u^2 + \frac{\partial^2 g}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't})h_1 h_2 uv + \frac{1}{2}\frac{\partial^2 g}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't})h_2^2 v^2 \\
&\quad - \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{jt}}h_1 u - \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{j't}}h_2 v \\
&\quad - \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}^2}h_1^2 u^2 - \frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}\partial z_{j't}}h_1 h_2 uv - \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{j't}^2}h_2^2 v^2] \\
&\quad \times k(u)k(v) du dv dz_{jt} dz_{j't},
\end{aligned}
$$

which can be simplified by the property of the kernel:

$$
\begin{aligned}
&\iiiint f^2 g g^+ du dv dz_{jt} dz_{j't} \\
&= \iiiint f^2(z_{jt}, z_{j't})[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})] \\
&\quad \times [0 + 0 + \frac{1}{2}\frac{\partial^2 g}{\partial z_{jt}^2}(\tilde{z}_{jt}, \tilde{z}_{j't}) h_1^2 u^2 + \frac{\partial^2 g}{\partial z_{jt}\partial z_{j't}}(\tilde{z}_{jt}, \tilde{z}_{j't}) h_1 h_2 uv + \frac{1}{2}\frac{\partial^2 g}{\partial z_{j't}^2}(\tilde{z}_{jt}, \tilde{z}_{j't}) h_2^2 v^2 \\
&\quad - \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{jt}} h_1 u - \frac{\partial \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})}{\partial z_{j't}} h_2 v \\
&\quad - \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}^2} h_1^2 u^2 - \frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{jt}\partial z_{j't}} h_1 h_2 uv - \frac{1}{2}\frac{\partial^2 \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid \tilde{z}_{jt}, \tilde{z}_{j't})}{\partial z_{j't}^2} h_2^2 v^2] \\
&\quad \times k(u) k(v) du dv dz_{jt} dz_{j't} \\
&= O(h_1^2) + O(h_2^2).
\end{aligned}
$$

where $O(h_1 h_2)$ disappears because the geometric mean is less than or equal to the arithmetic mean.

By the similar argument we have,

$$
\iiiint f f^+ g^2 du dv dz_{jt} dz_{j't} = O(h_1^2) + O(h_2^2)
$$

and

$$
\iiiint f f^+ g g^+ du dv dz_{jt} dz_{j't} = O(h_1^2) + O(h_2^2)
$$

Therefore, we conclude that

$$
\begin{aligned}
&h_1^{-1} h_2^{-1} \mathbb{E}\Big\{[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})] \\
&\quad \times [g(z_{j\tau}, z_{j'\tau}) - \mathbb{E}(\tilde{x}_\tau^\top \bar{\gamma}^u \mid z_{j\tau}, z_{j'\tau})] k\left(\frac{z_{jt} - z_{j\tau}}{h_1}\right) k\left(\frac{z_{j't} - z_{j'\tau}}{h_2}\right)\Big\} \\
&= \mathbb{E}\Big\{[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})]^2 f(z_{jt}, z_{j't})\Big\} + O(h_1^2) + O(h_2^2)
\end{aligned}
$$

□

Lemma D.9. *Under the alternative hypothesis,*

$$\begin{aligned}&\mathbb{P}(|\hat{\Phi} - \mathbb{E}\Big\{\big[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})\big]^2 f(z_{jt}, z_{j't})\Big\} - O(h_1^2) - O(h_2^2)| > \epsilon)\\ &\leq C_1 T^2 (e^{-C_2 h_1^2 h_2^2 N\epsilon^2/J\log^2 J} + e^{-C_2 N/J}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1^2 h_2^2 T\epsilon^2}.\end{aligned}$$

The problem is to choose the tuning parameters to separate the nulls from the alternatives. We use a theorem to conclude our arguments for selecting non-separating alternatives. Moreover, if the non-separating alternatives are well chosen, we can determine the preference parameters estimated by the 2SLS.

We rely on the two lemmas that for the null:

$$\mathbb{P}(|\hat{\Phi}| > \epsilon) \leq C_1 T^2 (e^{-C_2 h_1^2 h_2^2 N\epsilon^2/J\log^2 J} + e^{-C_2 N/J}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1 h_2 T\epsilon^2},$$

and for the alternative:

$$\begin{aligned}&\mathbb{P}(|\hat{\Phi} - \mathbb{E}\Big\{\big[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top \bar{\gamma}^u \mid z_{jt}, z_{j't})\big]^2 f(z_{jt}, z_{j't})\Big\} - O(h_1^2) - O(h_2^2)| > \epsilon)\\ &\leq C_1 T^2 (e^{-C_2 h_1^2 h_2^2 N\epsilon^2/J\log^2 J} + e^{-C_2 N/J}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1^2 h_2^2 T\epsilon^2}.\end{aligned}$$

Lemma D.10. *Assume* $\log J = o(h_1 h_2 T)$*,* $JT\log^2 J/h_1 h_2 = o(N)$*, and* $h_1 h_2 = o(1)$*. Under the null hypothesis,*

$$\mathbb{P}(|\hat{\Phi}| < c_0 \sqrt{\log(J \vee T)/h_1 h_2 T}) > 1 - c_1 (J \vee T)^{-c_2},$$

and under the alternative hypothesis,

$$\mathbb{P}(|\hat{\Phi}| > c_0 \sqrt{\log(J \vee T)/h_1 h_2 T}) > 1 - c_1 (J \vee T)^{-c_2}$$

for some absolute constant $c_0 > 0$*,* $c_1 > 0$*, and* $c_2 > 2$*.*

Proof. We begin with the null hypothesis:

$$\begin{aligned}\mathbb{P}(|\hat{\Phi}| > \epsilon) &\leq C_1 T^2 (e^{-C_2 h_1^2 h_2^2 N\epsilon^2/J\log^2 J} + e^{-C_2 N/J}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1 h_2 T\epsilon^2}\\ &\leq C_1 T^2 (e^{-C_2 h_1^2 h_2^2 JT\log^2 J\epsilon^2/h_1 h_2 J\log^2 J} + e^{-C_2 JT\log^2 J/h_1 h_2 J}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1 h_2 T\epsilon^2}\\ &\leq C_1 T^2 (e^{-C_2 h_1 h_2 T\epsilon^2} + e^{-C_2 T\log^2 J/h_1 h_2}) + \tilde{C}_1 e^{-\tilde{C}_2 h_1 h_2 T\epsilon^2}.\end{aligned}$$

Therefore,

$$
\begin{aligned}
\mathbb{P}(|\hat{\Phi}| > c_0\sqrt{\log(J\vee T)/h_1h_2T}) &\le C_1T^2(e^{-c_0^2C_2\log(J\vee T)} + e^{-C_2T\log^2 J/h_1h_2}) + \tilde{C}_1e^{-c_0^2\tilde{C}_2\log(J\vee T)}\\
&\le C_1T^2(e^{-c_0^2C_2\log(J\vee T)} + e^{-\tilde{c}_0C_2\log^2(J\vee T)}) + \tilde{C}_1e^{-c_0^2\tilde{C}_2\log(J\vee T)}\\
&\le \frac{C_1T^2}{(J\vee T)^{c_0^2C_2}} + \frac{C_1T^2}{(J\vee T)^{\tilde{c}_0C_2\log(J\vee T)}} + \frac{\tilde{C}_1}{(J\vee T)^{c_0^2C_2}}\\
&\le \frac{C_1T^2}{(J\vee T)^{c_0^2C_2}} + \frac{\tilde{c}C_1T^2}{(J\vee T)^{c_0^2C_2}} + \frac{\tilde{C}_1}{(J\vee T)^{c_0^2C_2}}\\
&\le c_{1,1}(J\vee T)^{-2}
\end{aligned}
$$

where $c_0 = 2\sqrt{\frac{1}{C_2}}$.

For the alternative hypothesis,

$$
\begin{aligned}
&\mathbb{P}(|\hat{\Phi} - \mathbb{E}\Big\{\big[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top\bar{\gamma}^u \mid z_{jt}, z_{j't})\big]^2 f(z_{jt}, z_{j't})\Big\} - O(h_1^2) - O(h_2^2)| > \epsilon)\\
&\le C_1T^2(e^{-C_2h_1^2h_2^2N\epsilon^2/J\log^2 J} + e^{-C_2N/J}) + \tilde{C}_1e^{-\tilde{C}_2h_1^2h_2^2T\epsilon^2}\\
&\le C_1T^2(e^{-\tilde{C}_2h_1h_2T\epsilon^2} + e^{-C_2JT\log^2 J/h_1h_2J}) + \tilde{C}_1e^{-\tilde{C}_2h_1^2h_2^2T\epsilon^2}
\end{aligned}
$$

For simplicity we denote $c_E := \mathbb{E}\Big\{\big[g(z_{jt}, z_{j't}) - \mathbb{E}(\tilde{x}_t^\top\bar{\gamma}^u \mid z_{jt}, z_{j't})\big]^2 f(z_{jt}, z_{j't})\Big\}$ we have

$$
\mathbb{P}(|\hat{\Phi} - c_E - O(h_1^2) - O(h_2^2)| < \frac{c_E}{2}) > 1 - [C_1T^2(e^{-4c_E^2\tilde{C}_2h_1h_2T} + e^{-C_2JT\log^2 J/h_1h_2J}) + \tilde{C}_1e^{-4c_E^2\tilde{C}_2h_1^2h_2^2T}]
$$

Since $c_E > 0$ by definition, we further have

$$
\begin{aligned}
\mathbb{P}(-\frac{c_E}{2} < \hat{\Phi} - c_E - O(h_1^2) - O(h_2^2) < \frac{c_E}{2}) > 1 - [C_1T^2(e^{-4c_E^2\tilde{C}_2h_1h_2T} + e^{-C_2JT\log^2 J/h_1h_2J}) + \tilde{C}_1e^{-4c_E^2\tilde{C}_2h_1^2h_2^2T}]\\
\mathbb{P}(\frac{c_E}{2} < \hat{\Phi} - O(h_1^2) - O(h_2^2) < \frac{3c_E}{2}) > 1 - [C_1T^2(e^{-4c_E^2\tilde{C}_2h_1h_2T} + e^{-C_2JT\log^2 J/h_1h_2J}) + \tilde{C}_1e^{-4c_E^2\tilde{C}_2h_1^2h_2^2T}]
\end{aligned}
$$

By a similar argument, we can have

$$
\mathbb{P}\left(\hat{\Phi} > \frac{c_E}{2} + O(h_1^2) + O(h_2^2)\right) > 1 - c_{2,1}(J\vee T)^{-2}
$$

Since $h_1 = o(1)$ and $h_2 = o(1)$, there exists T_E such that if $T > T_E$, $O(h_1^2) + O(h_2^2) > -\frac{c_s}{4}$. Therefore,

$$
\mathbb{P}\left(\hat{\Phi} > \frac{c_E}{4}\right) > 1 - c_{2,1}(J\vee T)^{-2}
$$

for $T > T_E$.

We want to show that there exists $\tilde{c}_{2,1}$ such that

$$\mathbb{P}\left(\hat{\Phi} > \frac{c_E}{4}\right) > 1 - \tilde{c}_{2,1}(J \vee T)^{-2}.$$

for any T.

With the fact that

$$\mathbb{P}\left(\hat{\Phi} > \frac{c_E}{4}\right) > 0$$

for $T < T_E$ where we also have $J < e^{T_E}$,

$$\mathbb{P}\left(\hat{\Phi} > \frac{c_E}{4}\right) > 1 - (e^{T_E} \vee T_E)^2(J \vee T)^{-2}.$$

Thus, $c_1 = \max\{c_{2,1}, (e^{T_E} \vee T_E)^2\}$ satisfies the conditions in the Lemma for the alternative hypothesis. □

This lemma suggests that increasing the number of markets is crucial for achieving sufficient power in the test. When setting the threshold, we also consider the number of alternatives; a higher threshold increases the likelihood that the test fails to reject the null hypothesis. Intuitively, it may seem that having more alternatives would strengthen the test. However, under the alternative hypothesis, despite this initial impression, an excessive number of alternatives does not improve the test's effectiveness because the bound for the alternative is not tight. The power of the test primarily comes from the number of markets. Assuming $\log J = o(T)$, adjusting the threshold by $\log J$ does not hinder the identification of the alternative hypothesis. Since $\log J = o(T)$, a large number of alternatives also requires a large number of markets. We include J in $\log(T \vee J)$ because if $\log J = o(T)$, incorporating $\log J$ into the test can help identify pure alternatives without significantly affecting the ability to distinguish non-pure alternatives. In the case of cross-validation, $\log J$ would be absorbed into the constant term.

Theorem 4. *(Pure Alternatives and Preference Parameters) Suppose Assumption 1 and 5 holds. Choose the tuning parameter* $\eta = c_0\sqrt{\log(J \vee T)/h_1 h_2 T}$ *for some* $c_0 > 0$. *With probability* $1 - c_1(J \vee T)^{-c_2}$ *for some* c_1>0 *and* c_2>0 *such that*

$$\hat{\mathcal{I}}_0 = \mathcal{I}_0$$

and for any $0 < \epsilon < 2$

$$\mathbb{P}(\|\hat{\gamma} - \gamma\|_2 > \epsilon) < C_1 T e^{-C_2 T \epsilon^2} + c_1(J \vee T)^{-c_2}.$$

for some C_1>0 and C_2>0.

Proof. For the first part of the theorem, when (j', j) is a pair of pure alternatives for the same type,

$$\mathbb{P}(|\hat{\Phi}| < c_0 \sqrt{\log(J \vee T)/h_1 h_2 T}) > 1 - c_1 (J \vee T)^{-\tilde{c}_2},$$

and when (j', j) is not a pair of pure alternatives for the same type,

$$\mathbb{P}(|\hat{\Phi}| > c_0 \sqrt{\log(J \vee T)/h_1 h_2 T}) > 1 - c_1 (J \vee T)^{-\tilde{c}_2}$$

for some absolute constant $c_0 > 0$, $c_1 > 0$, and $\tilde{c}_2 > 2$. This means that

$$\min\{\mathbb{P}(\text{selected} \mid \text{a pair of pures}), \mathbb{P}(\text{not selected} \mid \text{not a pair of pures})\} > 1 - c_1 (J \vee T)^{-\tilde{c}_2}.$$

We therefore have

$$\begin{aligned}
\mathbb{P}(\hat{\mathcal{I}}_0 = \mathcal{I}_0) &\geq \mathbb{P}(\frac{J(J-1)}{2} \text{ tests are correct}) \\
&\geq 1 - \frac{J(J-1)}{2} c_1 (J \vee T)^{-\tilde{c}_2} > 1 - c_1 (J \vee T)^{-c_2}
\end{aligned}$$

for some c_1>0 and c_2>0 by the union bound.

For the second part of the theorem, we know

$$\mathbb{P}\left(\|\hat{\gamma}^{jj'} - \gamma_k^K\|_2 > \epsilon; \exists k \ \ s.t. \ j, j' \in I(k)\right) < C_1 T e^{-C_2 T \epsilon^2}$$

for some positive C_1 and C_2 by Lemma 5. We can show that

$$\mathbb{P}\left(\|\hat{\gamma}^K - \gamma^K\|_2 > \epsilon \mid \hat{\mathcal{I}}_0 = \mathcal{I}_0\right) < C_1 T e^{-C_2 T \epsilon^2}$$

by the union bound. Then,

$$\begin{aligned}
\mathbb{P}\left(\|\hat{\gamma}^K - \gamma^K\|_2 > \epsilon\right) &= \mathbb{P}\left(\{\|\hat{\gamma}^K - \gamma^K\|_2 > \epsilon\} \cap \{\hat{\mathcal{I}}_0 = \mathcal{I}_0\}\right) + \mathbb{P}(\{\|\hat{\gamma}^K - \gamma^K\|_2 > \epsilon\} \cap \{\hat{\mathcal{I}}_0 \neq \mathcal{I}_0\}) \\
&\leq \mathbb{P}\left(\|\hat{\gamma}^K - \gamma^K\|_2 > \epsilon \mid \hat{\mathcal{I}}_0 = \mathcal{I}_0\right) \mathbb{P}\left(\hat{\mathcal{I}}_0 = \mathcal{I}_0\right) + \mathbb{P}\left(\hat{\mathcal{I}}_0 \neq \mathcal{I}_0\right) \\
&< C_1 T e^{-C_2 T \epsilon^2} + c_1 (J \vee T)^{-c_2}
\end{aligned}$$

□

E Fixed effects and probabilities of types

Not explicitly, we choose the tuning parameter as $\eta = c_0\sqrt{\log(J \vee T)/h_1h_2T}$ for some $c_0 > 0$ without being clear. This section discusses how to estimate the fixed effects. Define

$$\begin{aligned}
M(\tilde{e}) &= \mathbb{E}\left[\sigma_{jt}^{-1}(\tilde{e}, I_0)\hat{z}_{jt}^{\top}\right]\left[\hat{z}_{jt}\sigma_{jt}^{-1}(\tilde{e}, I_0)\right], \\
M_{JT}(\tilde{\gamma}, \tilde{e}) &= \sum_{j}\sum_{t}\sum_{\tau\neq t}\frac{1}{JT(T-1)}\sigma_{jt}^{-1}(\hat{I}_0)\hat{z}_{jt}^{\top}\hat{z}_{j\tau}\sigma_{j\tau}^{-1}(\tilde{e}, \hat{I}_0) \\
\hat{M}_{JT}(\tilde{e}) &= \sum_{j}\sum_{t}\sum_{\tau\neq t}\frac{1}{JT(T-1)}\hat{\sigma}_{jt}^{-1}(\tilde{e}, \hat{I}_0)\hat{z}_{jt}^{\top}\hat{z}_{j\tau}\hat{\sigma}_{j\tau}^{-1}(\tilde{e}, \hat{I}_0)
\end{aligned}$$

where $\hat{s}$ and $\hat{\gamma}$. We also have

$$\hat{e} = \operatorname{argmin}_{\tilde{e}}\hat{M}_{JT}(\tilde{e})$$

and we want to show that $M(\hat{e})$ is close to $M(e^{I_0})$. By definition of the minimization problem, we notice $\hat{M}_{JT}(\hat{e}) \leq \hat{M}_{JT}(e^{I_0})$ and $M(e^{I_0}) \leq \hat{M}_{JT}(\hat{e})$, so

$$\begin{aligned}
|M(\hat{e}) - M(e^{I_0})| &\leq |\hat{M}_{JT}(\hat{e}) - M(\hat{e})| + |\hat{M}_{JT}(e^{I_0}) - M(e^{I_0})| \\
&\leq |\hat{M}_{JT}(\hat{e}) - M_{JT}(\hat{e})| + |M_{JT}(\hat{e}) - M(\hat{e})| + |\hat{M}_{JT}(e^{I_0}) - M_{JT}(e^{I_0})| + |M_{JT}(e^{I_0}) - M(e^{I_0})|
\end{aligned}$$

where the second inequality is the triangular inequality.

Equation (13) gives $\xi_{jt} = \hat{\sigma}_{jt}^{-1}(\tilde{e}_{\bullet j}, \tilde{\xi}_t^K)$ from the inverse function:

$$\hat{s}_{jt} = \sum_{k}\frac{\tilde{e}_{kj}\hat{s}_{j_kt}e^{\lambda_k\xi_{jt}+x_{jt}^{\top}\hat{\beta}_k-\hat{\alpha}_kp_{j_kt}}}{e^{\tilde{\xi}_{j_kt}+x_{jt}^{\top}\hat{\beta}_k-\hat{\alpha}_kp_{j_kt}}} \tag{13}$$

$$\sup_{e\in E}|\hat{M}_{JT}(e) - M_{JT}(e)| < m_1\|\log\hat{S} - \log S\|_{\max} + m_2\|\hat{\gamma}^K - \gamma^K\|_2$$

Lemma E.1. *Under assumption, for any $\epsilon > 0$,*

$$\mathbb{P}(|M(\hat{e}) - M(e^{I_0})| > \epsilon) < C_1Te^{-C_2T\epsilon^2} + c_1(J \vee T)^{-c_2}$$

for some constants, $c_1 > 0$, $c_2 > 0$, $C_1 > 0$ and $C_2 > 0$.

Proof. By the union bounds,

$$
\begin{aligned}
\mathbb{P}(|M(\hat{e}) - M(e^{I_0})| > \epsilon) < \ & \mathbb{P}(|\hat{M}_{JT}(\hat{e}) - M_{JT}(\hat{e})| > \frac{\epsilon}{4}) + \mathbb{P}(|\hat{M}_{JT}(e^{I_0}) - M_{JT}(e^{I_0})| > \frac{\epsilon}{4}) \\
& + \mathbb{P}(|M_{JT}(\hat{e}) - M(\hat{e})| > \frac{\epsilon}{4}) + \mathbb{P}(|M_{JT}(e^{I_0}) - M(e^{I_0})| > \frac{\epsilon}{4}) \\
& \leq 2\mathbb{P}(\sup_e |\hat{M}_{JT}(e) - M(e)| > \frac{\epsilon}{4}) + 2\mathbb{P}(\sup_e |M_{JT}(e) - M(e)| > \frac{\epsilon}{4}) \\
& \leq 2 \sup_e \mathbb{P}(|\hat{M}_{JT}(e) - M(e)| > \frac{\epsilon}{4}) + 2 \sup_e \mathbb{P}(|M_{JT}(e) - M(e)| > \frac{\epsilon}{4}) \\
& \leq C_1 e^{-C_2 T \epsilon^2} + c_1 (J \vee T)^{-c_2}
\end{aligned}
$$

□

Theorem 5. *Under Assumptions 1, 5, and 6, for any* $\epsilon > 0$*,*

$$
\mathbb{P}(\|\hat{e} - e^{I_0}\|_2^2 > \epsilon) < C_1 e^{-C_2 T \epsilon^2 / J^2} + c_1 (J \vee T)^{-c_2},
$$

for some constant $C_1 > 0$ *and* $C_2 > 0$*. And*

$$
\mathbb{P}(|\hat{\xi}_{jt} - \xi_{jt}| > \epsilon) < C_1 e^{-C_2 T \epsilon^2 / J^2} + C_3 e^{-C_4 J \epsilon^2} + c_1 (J \vee T)^{-c_2},
$$

for some constants $C_1 > 0, C_2 > 0, C_3 > 0$*, and* $C_4 > 0$*.*

Proof. Using Taylor expansion on the moment function, we have,

$$
\begin{aligned}
\mathbb{P}(|M(\hat{e}) - M(e^{I_0})| > \epsilon) &= \mathbb{P}(\frac{1}{2}(\hat{e} - e^{I_0})^\top |H_M(e)| (\hat{e} - e^{I_0}) > \epsilon) \\
&> \mathbb{P}(\frac{C_3 \|\hat{e} - e^{I_0}\|_2^2}{J} > \epsilon)
\end{aligned}
$$

Therefore, we have

$$
\mathbb{P}(\|\hat{e} - e^{I_0}\|_2^2 > \frac{J\epsilon}{C_3}) < C_1 e^{-C_2 T \epsilon^2} \Rightarrow \mathbb{P}(\|\hat{e} - e^{I_0}\|_2^2 > \epsilon) < C_1 e^{-\tilde{C}_2 T \epsilon^2 / J^2}.
$$

Define

$$
\begin{aligned}
\varsigma_k(\tilde{e}, \tilde{\xi}_t^K) &= \mathbb{E}\left[\sigma_{jt}^{-1}(\tilde{e}, I_0, \tilde{\xi}_t^K) \mid j \notin J - I(k)\right], \\
\varsigma_{kJ}(\tilde{e}) &= \sum_{j \notin J - I(k)} \sigma_{jt}^{-1}(\tilde{e}, I_0, \tilde{\xi}_t^K), \\
\hat{\varsigma}_{kJ}(\tilde{e}, \tilde{\xi}_t^K) &= \sum_{j \notin J - I(k)} \hat{\sigma}_{jt}^{-1}(\tilde{e}, I_0, \tilde{\xi}_t^K).
\end{aligned}
$$

Similarly, we can show that

$$\mathbb{P}(|\hat{\xi}_{jt} - \xi_{jt}| > \epsilon) < C_1 e^{-C_2 T \epsilon^2 / J^2} + C_3 e^{-C_4 J \epsilon^2}.$$

□

Theorem 6. *Under Assumptions 1, 5, and 6, for any* $\epsilon > 0$*,*

$$\mathbb{P}(\|\hat{\pi}_{kt} - \pi_{kt}\|_2^2 > \epsilon) < C_1 e^{-C_2 T \epsilon^2 / J^2} + C_3 e^{-C_4 J \epsilon^2} + c_1 (J \vee T)^{-c_2}.$$

for some constants $C_1 > 0, C_2 > 0, C_3 > 0$*, and* $C_4 > 0$*.*

Proof. We aim to bound the difference between π_{kt} and $\hat{\pi}_{kt}$. The expressions for π_{kt} and $\hat{\pi}_{kt}$ are

$$\pi_{kt} = \frac{e^{\lambda_k \xi_{j_k t} + x_{j_k t}^\top \beta_k - \alpha_k p_{j_k t}}}{s_{j_k t} \sum_{i \in J} e_{ki}^{I_0} e^{\xi_{it} + x_{it}^\top \beta_k - \alpha_k p_{it}}},$$
$$\hat{\pi}_{kt} = \frac{e^{\lambda_k \hat{\xi}_{j_k t} + x_{j_k t}^\top \hat{\beta}_k - \hat{\alpha}_k p_{j_k t}}}{s_{j_k t} \sum_{i \in J} \hat{e}_{ki}^{I_0} e^{\xi_{it} + x_{it}^\top \hat{\beta}_k - \hat{\alpha}_k p_{it}}}.$$

The denominators for π_{kt} and $\hat{\pi}_{kt}$ are

$$D = s_{j_k t} \sum_{i \in J} e_{ki}^{I_0} e^{\xi_{it} + x_{it}^\top \beta_k - \alpha_k p_{it}},$$
$$\hat{D} = s_{j_k t} \sum_{i \in J} \hat{e}_{ki}^{I_0} e^{\xi_{it} + x_{it}^\top \hat{\beta}_k - \hat{\alpha}_k p_{it}}.$$

The difference between the two denominators $|D - \hat{D}|$ can be expanded as

$$\begin{aligned} |D - \hat{D}| &= \left| s_{j_k t} \sum_{i \in J} \left(e_{ki}^{I_0} e^{\xi_{it} + x_{it}^\top \beta_k - \alpha_k p_{it}} - \hat{e}_{ki}^{I_0} e^{\xi_{it} + x_{it}^\top \hat{\beta}_k - \hat{\alpha}_k p_{it}} \right) \right| \\ &\leq s_{j_k t} \sum_{i \in J} e_{ki}^{I_0} e^{\xi_{it} + x_{it}^\top \beta_k - \alpha_k p_{it}} \left(|\hat{e}_{ki}^{I_0} - e_{ki}^{I_0}| + |x_{it}^\top (\hat{\beta}_k - \beta_k)| + |(\hat{\alpha}_k - \alpha_k) p_{it}| \right). \end{aligned}$$

We now consider the difference between the inverses of the denominators. Using the mean value theorem for inverses, we have

$$\left| \frac{1}{D} - \frac{1}{\hat{D}} \right| = \frac{|D - \hat{D}|}{|D||\hat{D}|}.$$

Specifically, we assume $D, \hat{D} \geq C_D > 0$ where C_D is a positive constant. Therefore, the inverse difference is bounded as:

$$\left|\frac{1}{D} - \frac{1}{\hat{D}}\right| \leq \frac{|D - \hat{D}|}{C_D^2}.$$

Now, combining the bounds on the numerator and the inverse of the denominator, we obtain the total bound for $|\pi_{kt} - \hat{\pi}_{kt}|$:

$$|\pi_{kt} - \hat{\pi}_{kt}| \leq \frac{|F - \hat{F}|}{C_D} + \frac{F}{C_D^2}|D - \hat{D}|,$$

where F and $\hat{F}$ are the numerators of π_{kt} and $\hat{\pi}_{kt}$, respectively.

Given the bounds on the numerator and denominator differences, we conclude:

$$|\pi_{kt} - \hat{\pi}_{kt}| \leq C\left(|\hat{\xi}_{j_kt} - \xi_{j_kt}| + |x_{j_kt}^\top(\hat{\beta}_k - \beta_k)| + |(\hat{\alpha}_k - \alpha_k)p_{j_kt}| + \max_{i\in J}|\hat{e}_{ki}^{I_0} - e_{ki}^{I_0}|\right),$$

where C is a constant that depends on the lower bound C_D for the denominators.

$$\begin{aligned}\mathbb{P}(|\pi_{kt} - \hat{\pi}_{kt}| > \epsilon) &\leq \mathbb{P}\left[C\left(|\hat{\xi}_{j_kt} - \xi_{j_kt}| + \|\hat{\beta}_k - \beta_k\|_2 + |\hat{\alpha}_k - \alpha_k| + \max_{i\in J}|\hat{e}_{ki}^{I_0} - e_{ki}^{I_0}| > \epsilon\right)\right] \\ &\leq C_1 e^{-C_2 T\epsilon^2/J^2} + C_3 e^{-C_4 J\epsilon^2} + c_1(J \vee T)^{-c_2}.\end{aligned}$$

□

F Tools and properties in high dimensions

F.1 Concentration inequalities, union bounds, and U statistics

Lemma F.1. *(Markov's Inequality) If X is a nonnegative random variable and $\epsilon > 0$, then the probability that X is at least ϵ is at most the expectation of X divided by ϵ*

$$\mathbb{P}(X \geq \epsilon) \leq \frac{\mathbb{E}(X)}{\epsilon}.$$

Lemma F.2. *(Hoeffding's Inequality) Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $X_i \in [a_i, b_i]$ for $i \in [n]$. Let $S_n = \sum_{i=1}^n X_i$. Then, for any $\epsilon > 0$,*

$$\mathbb{P}\left(S_n - \mathbb{E}(S_n) \geq \epsilon\right) \leq e^{-2\epsilon^2/\sum_{i=1}^n(b_i - a_i)^2}, \qquad \mathbb{P}\left(\mathbb{E}(S_n) - S_n \geq \epsilon\right) \leq e^{-2\epsilon^2/\sum_{i=1}^n(b_i - a_i)^2}.$$

and

$$\mathbb{P}\left(|S_n - \mathbb{E}(S_n)| \geq \epsilon\right) \leq 2e^{-2\epsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2}.$$

Definition 4. *(Sub-Gaussian) A random variable X is called **sub-Gaussian** if its tail probabilities decay at least as fast as those of a Gaussian distribution. Specifically, X is sub-Gaussian if there exists a positive constant $\sigma^2 > 0$ such that for all $\epsilon > 0$,*

$$\mathbb{P}(|X| > \epsilon) \leq 2e^{-\epsilon^2/2\sigma^2}.$$

An equivalent definition involves the moment-generating function (MGF). A random variable X is sub-Gaussian if there exists a constant $\sigma^2 > 0$ such that for all $\lambda \in \mathbb{R}$,

$$\mathbb{E}(e^{\lambda X}) \leq e^{\lambda^2\sigma^2/2}.$$

In both definitions, σ^2 is often referred to as the sub-Gaussian parameter and is similar to the variance in the Gaussian distribution, controlling the rate of tail decay and the growth of the MGF.

Lemma F.3. *(Bounded Random Variable) Let X be a random variable such that $a \leq X \leq b$ almost surely, where a and b are constants. Then X is sub-Gaussian with variance proxy parameter $\sigma^2 = \frac{(b-a)^2}{4}$.*

Lemma F.4. *(Product of sub-Gaussian Variable) Let X be a bounded random variable such that $|X| \leq M$ almost surely, for some $M > 0$, and let Y be a sub-Gaussian random variable with parameter σ. Then, the product $Z = XY$ is sub-Gaussian with a parameter at most $M\sigma$.*

Proof. By definition, X is sub-Gaussian if there exists a constant $\sigma^2 > 0$ such that for all $\lambda \in \mathbb{R}$ $\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\lambda^2\sigma^2/2}$. Consider the MGF of $Z = YX$, since $|Y| \leq M$, it follows that

$$\mathbb{E}(e^{\lambda YX}) \leq \mathbb{E}(e^{\lambda M|X|}) \leq e^{\lambda^2 M^2\sigma^2/2},$$

by using the sub-Gaussian property of $|X|$ (it is sub-Gaussian by the first definition), therefore

$$\mathbb{E}(e^{\lambda Z}) \leq e^{\lambda^2 M^2\sigma^2/2}.$$

The inequality above shows that XY is sub-Gaussian with parameter $M^2\sigma^2$. □

Lemma F.5. *(Union Bound) Let $A_1, A_2, \ldots, A_n$ be events in a probability space. Then, the probability that at least one of these events occurs is bounded above by the sum of their probabilities.*

Formally, the union bound is given by

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

The following lemmas are standard in the literature and are presented here to streamline our subsequent proofs of estimates. They also offer insights into how to bound tail probabilities.

Lemma F.6. *(Union Bound Variation) Let X and Y be random variables. Then, for any $\epsilon > 0$, the following bound holds:*

$$\mathbb{P}(|X||Y| > \epsilon) \leq \mathbb{P}(|X| > \sqrt{\epsilon}) + \mathbb{P}(|Y| > \sqrt{\epsilon}).$$

Proof. The event $|X||Y| > \epsilon$ implies that either $|X|$ or $|Y|$ must be sufficiently large for their product to exceed ϵ. More precisely, the event $\{|X||Y| > \epsilon\}$ can be rewritten as

$$\{|X||Y| > \epsilon\} \subseteq \{|X| > \sqrt{\epsilon}\} \cup \{|Y| > \sqrt{\epsilon}\}.$$

which means that at least one of the events $\{|X| > \sqrt{\epsilon}\}$ and $\{|Y| > \sqrt{\epsilon}\}$ happens. Using the union bound, we have

$$P(|X||Y| > \epsilon) \leq P(|X| > \sqrt{\epsilon}) + P(|Y| > \sqrt{\epsilon}).$$

□

Lemma F.7. *(Union Bound Variation) Given random variables X^l for $l \in [L] > 0$, assume there exists functions f_l for any $\epsilon > 0, \mathbb{P}(X^l > \epsilon) < g_l(\epsilon)$. Then, that the probability of the summation of the random variables is at least ϵ can be bounded as:*

$$\mathbb{P}(\sum_{l \in [L]} X^l > \epsilon) < \sum_{l \in [L]} g_l(\frac{\epsilon}{L}).$$

Proof. We notice that

$$\mathbb{P}(\sum_{l \in [L]} X^l > \epsilon) \leq \mathbb{P}(\max_{l \in [L]} X^l > \frac{\epsilon}{L}).$$

This is because the set on the equation's left is a subset of the set on the right; otherwise, if all $X^l \leq \epsilon/L$, the summation will be smaller than L.

Then we have,

$$\mathbb{P}(\max_{l\in[L]} X^l > \epsilon/L) = \mathbb{P}(\bigcup_{l\in[L]}\{X^l > \epsilon/L\}) \leq \sum_{l\in[L]} \mathbb{P}(X^l > \epsilon/L) < \sum_{l\in[L]} g_l(\frac{\epsilon}{L}),$$

where the first inequality is from the Lemma F.5 of the union bound, and the second inequality holds by assumption. □

Lemma F.8. *(Constant in Bound) Given a random variable X for $n \geq 1$, assume there exists function g for any $\epsilon > 0$, $\mathbb{P}(X > \epsilon) < g(\epsilon)$. Then for any $c > 0$*

$$\mathbb{P}(cX_n > \epsilon) < g(\frac{\epsilon}{c}).$$

Proof. The proof is direct by noticing

$$\mathbb{P}(cX > \epsilon) = \mathbb{P}(X > \frac{\epsilon}{c}).$$

□

Lemma F.9. *(Bound of Average) Given random variables X^l for $l \in [L] > 0$, assume there exists functions g_l where for any $\epsilon > 0, \mathbb{P}(X^l > \epsilon) < g_l(\epsilon)$. Then, we have the bound for the average of these random variables:*

$$\mathbb{P}(\frac{1}{L}\sum_{l\in[L]} X^l > \epsilon) < \sum_{l\in[L]} g_l(\epsilon).$$

Proof. This comes directly from Lemma F.7 and Lemma F.8. □

Lemma F.10. *(Upper-bound of Bound) Given random variables X and Y and assume $Y \geq X$ almost surely, then for any constant ϵ*

$$\mathbb{P}(X > \epsilon) \leq \mathbb{P}(Y > \epsilon)$$

Proof. This is because if an event satisfies $X > \epsilon$ on the probability space, the event satisfies $Y > \epsilon$ by the assumption that $Y \geq X$. □

The Lemma F.10 implies that we can use the tail bound of Y to derive the tail bound for X. By contrast, we also have $\mathbb{P}(X > \epsilon_1) \geq \mathbb{P}(X > \epsilon_2)$ for $\epsilon_1 \leq \epsilon_2$ to derive the bound, which comes from the definition of the CDF. We use these facts regularly without explicitly mentioning them throughout the proofs.

We nonparametrically test the form of the function where U Statistics are helpful.

Definition 5. *(U Statistics) A U-statistic is defined for a set of T observations $X_1, X_2, \ldots, X_T$ and is of the form:*

$$U_T = \frac{1}{\binom{T}{k}} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le T} h(X_{i_1}, X_{i_2}, \ldots, X_{i_k}),$$

where $h(\cdot)$ is a symmetric kernel function of k variables (the order of the U-statistic), meaning that the function does not change if the inputs are permuted, and $k \le T$.

Lemma F.11. *(Hoeffding's Inequality for U-Statistics) Let U_T be a U-statistic of order k with kernel $h(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$, where $X_1, X_2, \ldots, X_T$ are independent random variables, and assume that the kernel h is bounded by*

$$|h(X_{i_1}, X_{i_2}, \ldots, X_{i_k})| \le M.$$

Then, Hoeffding's inequality for U-statistics is given by

$$\mathbb{P}\left(U_T - \mathbb{E}[U_T] \ge \epsilon\right) \le 2e^{-T\epsilon^2/2kM^2},$$

where ϵ is the deviation threshold.

Hoeffding's inequality for the U statistics is most commonly used because of its simplicity. However, like Lemma F.15, when the numerator in the power of the exponential is tiny, the bound does not perform well. Specifically, when the ϵ is small, the power of the square on it will require the number of observations T to be large—the recent development of Bernstein's version of the concentration parameter deals with this problem.

Lemma F.12. *(Bernstein's Inequality for U-Statistics) Let U_T be a U-statistic of order k with kernel $h(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$, where $X_1, X_2, \ldots, X_T$ are independent random variables, and assume that the kernel h is bounded by*

$$|h(X_{i_1}, X_{i_2}, \ldots, X_{i_k})| \le M.$$

Then Bernstein's inequality for U-statistics is given by

$$\mathbb{P}\left(|U_T - \mathbb{E}[U_T]| > \epsilon\right) \le 2e^{-T\epsilon^2/(4k\sigma^2 + \frac{2}{3}kM\epsilon)},$$

where σ^2 is the variance of $h(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$.

This inequality is useful when we estimate the properties of an estimator condition on events.

We conclude with another commonly used fact which is helpful for our discussion of the 2SLS later.

Lemma F.13. *Let $X \in \mathbb{R}^{T\times p}$ and $Z \in \mathbb{R}^{T\times q}$ be matrices such that (a) The eigenvalues of $\frac{1}{T}Z^\top Z$ are strictly positive. (b) The singular values of $\frac{1}{T}Z^\top X$ are strictly positive. Then, the eigenvalues of the matrix $\frac{1}{T}X^\top Z(Z^\top Z)^{-1}Z^\top X$ are strictly positive.*

Proof. Let $A = \frac{1}{T}X^\top Z(Z^\top Z)^{-1}Z^\top X$. We aim to show that the eigenvalues of A are bounded from below. Consider the quadratic form of A,

$$v^\top Av = \frac{1}{T}v^\top X^\top Z(Z^\top Z)^{-1}Z^\top Xv = \frac{1}{T}w^\top (Z^\top Z)^{-1}w,$$

where $w = Z^\top Xv$. Let $Z^\top Z = U\Lambda U^\top$, where U is an orthogonal matrix (containing the eigenvectors of $Z^\top Z$) and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_q)$ is a diagonal matrix with the eigenvalues of $Z^\top Z$, denoted $\lambda_1, \lambda_2, \dots, \lambda_q$, which are all strictly positive.

Thus, $(Z^\top Z)^{-1} = U\Lambda^{-1}U^\top$, where $\Lambda^{-1} = \mathrm{diag}(\lambda_1^{-1}, \dots, \lambda_q^{-1})$.

Now, we transform the quadratic form:

$$w^\top (Z^\top Z)^{-1}w = (U^\top w)^\top \Lambda^{-1}(U^\top w). \tag{14}$$

Let $w' = U^\top w$, and hence the quadratic form becomes:

$$w^\top (Z^\top Z)^{-1}w = \sum_{i=1}^{q} \lambda_i^{-1} w_i'^2 \geq \frac{1}{\lambda_{\min}}\|w'\|_2^2, \tag{15}$$

where w_i' are the components of w' and the last inequality is from the Cauchy Schwarz (CS) inequality. In the matrix format, this is equivalent to

$$w^\top (Z^\top Z)^{-1}w = \frac{1}{T}w^\top (\frac{Z^\top Z}{T})^{-1}w \geq \frac{\|w\|_2^2}{T\lambda_{\min}\left(\frac{Z^\top Z}{T}\right)},$$

Given the definition of w,

$$\|w\|_2^2 = v^\top X^\top ZZ^\top Xv \geq T^2\sigma_{\min}^2\|v\|_2^2,$$

where $\sigma_{\min}$ is the smallest singular value of $\frac{1}{T}X^\top Z$. Then we have

$$w^\top (Z^\top Z)^{-1}w \geq \frac{T\sigma_{\min}^2\|v\|^2}{\lambda_{\min}\left(\frac{Z^\top Z}{T}\right)}.$$

Thus, by the property of Rayleigh Quotient, the lower bound on the eigenvalues of A is

$$\lambda(A) \geq \frac{\sigma^2_{\min}}{\lambda_{\min}\left(\frac{Z'Z}{T}\right)} > 0.$$

□

F.2 Non-asymptotic properties for the empirical shares

Given alternative j and market t, the observed sale is Y_{jt}, we have that $Y_{jt} \sim \text{Binomial}(N, s_{jt})$ and $\hat{s}_{jt} = \frac{Y_{jt}}{N}$.

Lemma F.14. *(Bound of Share) If all parameter spaces are compact,* $\min_{k\in[K],t\in[T]} \pi_{kt} > \varphi_\pi$ *for some* $\varphi_\pi > 0$*, and* $\min_{k\in[K]} \frac{[J_k]}{[J]} > \varphi_J$ *for some* $\varphi_J > 0$*, then there exist constants* $\varphi^+ > 0$ *and* $\varphi^- > 0$ *such that*

$$\frac{\varphi^-}{J} < \max_{j\in[J],t\in[T]} s_{jt} < \frac{\varphi^+}{J}.$$

Proof. We use constants $\varphi^-_\delta < \delta_{kjt} < \varphi^+_\delta$ to reflect that δ_{kjt} is uniformly bounded for any type k, alternative j, and market t, since it is from the linear combination of bounded variables. We have

$$s_{jt} = \sum_{k\in[K]} \pi_{kt} \frac{\mathbb{1}_{J_k}(j) e^{\delta_{kjt}}}{\sum_{i\in J_k} e^{\delta_{kit}}}$$

by definition.

We also have

$$\sum_{k\in[K]} \pi_{kt} \frac{\mathbb{1}_{J_k}(j) e^{\delta_{kjt}}}{\sum_{i\in J_k} e^{\delta_{kit}}} \leq \sum_{k\in[K]} \pi_{kt} \frac{e^{\varphi^+_\delta}}{\sum_{i\in J_k} e^{\varphi^-_\delta}} = \frac{e^{\varphi^+_\delta}}{\sum_{i\in J_k} e^{\varphi^-_\delta}} \leq \frac{e^{\varphi^+_\delta}}{\varphi_J J e^{\varphi^-_\delta}} < \frac{\varphi^+}{J}$$

where the second inequality comes from the sum of π_{kt} equals to 1.

On the other side,

$$\sum_{k\in[K]} \pi_{kt} \frac{\mathbb{1}_{J_k}(j) e^{\delta_{kjt}}}{\sum_{i\in J_k} e^{\delta_{kit}}} \geq \sum_{k\in[K]} \varphi_\pi \frac{e^{\varphi^-_\delta}}{\sum_{i\in J_k} e^{\varphi^+_\delta}} \geq \frac{K\varphi_\pi e^{\varphi^-_\delta}}{\sum_{i\in J_k} e^{\varphi^+_\delta}} \geq \frac{K\varphi_\pi e^{\varphi^-_\delta}}{J e^{\varphi^+_\delta}} > \frac{\varphi^-}{J}.$$

□

With these two bounds, we can get some practical properties to observe that the empirical shares are more significant than specific values with high probabilities.

Lemma F.15. *(Hoeffding's Threshold for Empirical Share) Given alternative $j \in [J]$ and market $t \in [T]$, the probability that $\hat{s}_{jt}$ exceeds a positive value φ where $\varphi < s_{jt}$ can be bounded as follows:*

$$\mathbb{P}\left(\hat{s}_{jt} \geq \varphi\right) \geq 1 - e^{-2N(s_{jt}-\varphi)^2}.$$

Proof. The proof follows from Lemma F.2 of Hoeffding's Inequality. We notice

$$\begin{aligned}\mathbb{P}\left(\hat{s}_{jt} > \varphi\right) &= 1 - \mathbb{P}(\hat{s}_{jt} \leq \varphi) = 1 - \mathbb{P}\left(Y_{jt} \leq N\varphi\right) = 1 - \mathbb{P}\left[Ns_{jt} - Y_{jt} \geq N(s_{jt} - \varphi)\right] \\ &\geq 1 - e^{-2N(s_{jt}-\varphi)^2},\end{aligned}$$

where in the last inequality, we use the property of binomial that $\mathbb{E}(Y_{jt}) = Ns_{jt}$ and $\epsilon = s_{jt} - \varphi > 0$. □

Lemma F.16. *(Hoeffding's Bound for Empirical Shares) Given alternative $j \in [J]$ and market $t \in [T]$, for any $\epsilon > 0$ the probability*

$$\mathbb{P}(|\hat{s}_{jt} - s_{jt}| \geq \epsilon) \leq 2e^{-2N\epsilon^2}$$

Proof. We use the Hoeffding's inequality,

$$\mathbb{P}(|\hat{s}_{jt} - s_{jt}| \geq \epsilon) = \mathbb{P}(N|\hat{s}_{jt} - s_{jt}| \geq N\epsilon) = \mathbb{P}(|Y_{jt} - Ns_{jt}| \geq N\epsilon) \leq 2e^{-2N\epsilon^2}$$

because Y_{jt} is the summation of the independent Bernoulli trials and $\mathbb{E}(Y_{jt}) = Ns_{jt}$. □

Hoeffding uses boundedness and independence. s_{jt} or ϵ appears in the power of square in the bound; therefore, it is not very good when s_{jt} is small. We can use the Chernoff bound to get a tighter bound of a binomial distribution when s_{jt} or ϵ is small.

Lemma F.17. *(Chernoff's Bound for Binomial Distribution) Given alternative $j \in [J]$ and market $t \in [T]$, for any $0 < \kappa < 1$, the following bounds hold:*

$$\begin{aligned}\textbf{\textit{Upper tail bound:}} \quad & \mathbb{P}[Y_{jt} \geq Ns_{jt}(1+\kappa)] \leq e^{-\kappa^2 Ns_{jt}/3} \\ \textbf{\textit{Lower tail bound:}} \quad & \mathbb{P}[Y_{jt} \leq Ns_{jt}(1-\kappa)] \leq e^{-\kappa^2 Ns_{jt}/2}\end{aligned}$$

Lemma F.18. *(Chernoff's Threshold for Empirical Share) Given alternative $j \in [J]$ and market $t \in [T]$, the probability that $\hat{s}_{jt}$ exceeds a positive value φ where $\varphi < s_{jt}$ can be bounded as* *follows:*

$$\mathbb{P}(\hat{s}_{jt} > \varphi) > 1 - e^{-N(s_{jt}-\varphi)^2/2s_{jt}}.$$

Proof. We set $\kappa = \frac{s_{jt}-\varphi}{s_{jt}}$ and thus:

$$\mathbb{P}(\hat{s}_{jt} \leq \varphi) \leq e^{-N(s_{jt}-\varphi)^2/2s_{jt}}.$$

for the Chernoff lower bound. We can choose $\kappa = \frac{s_{jt}-\varphi}{s_{jt}}$ since $\frac{s_{jt}-\varphi}{s_{jt}} > 0$ due to the fact that $s_{jt} - \varphi > 0$ and $s_{jt} > 0$. $\frac{s_{jt}-\varphi}{s_{jt}} < 1$ is by $1 - \frac{\varphi}{s_{jt}} < 1$ from $\varphi > 0$. □

Lemma F.19. *(Chernoff's Bound for Empirical Share) Given alternative $j \in [J]$ and market $t \in [T]$, for any $0 < \epsilon < s_{jt}$ the probability*

$$\mathbb{P}(|\hat{s}_{jt} - s_{jt}| \geq \epsilon) \leq e^{-N\epsilon^2/3s_{jt}}$$

Proof. Similarly, let $\kappa = \frac{\epsilon}{s_{jt}}$

$$\mathbb{P}(|\hat{s}_{jt} - s_{jt}| \geq \epsilon) \leq e^{-N\epsilon^2/3s_{jt}}.$$

□

Lemma F.20. *Given alternative $j \in [J]$ and market $t \in [T]$, if $\frac{\varphi_0}{J} < s_{jt} < \frac{\varphi_1}{J}$, we have a high probability that the empirical share $\hat{s}_{jt}$ is larger than $\frac{\varphi_0}{2J}$—that is*

$$\mathbb{P}(\hat{s}_{jt} > \frac{\varphi_0}{2J}) > 1 - e^{-N\varphi_0^2/8c_1 J}$$

which is equivalent to

$$\mathbb{P}(\hat{s}_{jt} \leq \frac{\varphi_0}{2J}) \leq e^{-N\varphi_0^2/8c_1 J}.$$

Proof. This comes directly from Lemma F.18 by setting $\varphi = \frac{\varphi_0}{2J}$. To be specific,

$$\mathbb{P}(\hat{s}_{jt} \leq \frac{\varphi_0}{2J}) \leq e^{-N(s_{jt}-\varphi_0/2J)^2/2s_{jt}} \leq e^{-N\varphi_0^2/8J^2 s_{jt}} \leq e^{-N\varphi_0^2/8\varphi_1 J}$$

where the second inequality is from the fact that s_{jt} is bounded below by $\frac{\varphi_0}{J}$ and the third inequality comes from fact that s_{jt} is bounded above by $\frac{\varphi_1}{J}$. □

The following sequence of lemmas shows that after dropping some small shares, the value

$|\mathbb{1}\{\hat{s}_{jt} > \delta\} \log \hat{s}_{jt} - \log s_{jt}|$ will be close to zero with high probability when the number of consumers is relatively large—not rigorously speaking when $N >> J$.

The distance of $\mathbb{1}\{\hat{s}_{jt} > \delta\} \log \hat{s}_{jt}$ and $\log s_{jt}$ can be decomposed into two parts $|\mathbb{1}\{\hat{s}_{jt} > \delta\}(\log \hat{s}_{jt} - \log s_{jt})| + |\mathbb{1}\{\hat{s}_{jt} \leq \delta\} \log s_{jt}|$. We deal with $|\mathbb{1}\{\hat{s}_{jt} > \delta\}(\log \hat{s}_{jt} - \log s_{jt})|$ first. The idea is to observe that when $\hat{s}_{jt}$ is not close to zero, we can check the distance of logarithmic distance by $\hat{s}_{jt} - s_{jt}$ directly. Meanwhile, the probability that $\hat{s}_{jt}$ close to zero is small. The tuning parameter is to ensure that the logarithms of shares are well-defined.

Lemma F.21. *Given alternative $j \in [J]$ in market $t \in [T]$, assume utilities are bounded such that $\frac{\varphi^-}{J} < s_{jt} < \frac{\varphi^+}{J}$. Given an alternative j in market t, for any $0 < \epsilon < 2$,*

$$\mathbb{P}[|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon] < C_1(e^{-C_2 N\epsilon^2/J} + e^{-C_2 N/J}).$$

where C_1 and C_2 are positive constants.

Proof. We decompose $\mathbb{P}[|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon]$ conditional on two events $\{\hat{s}_{jt} > \frac{\varphi^-}{2J}\}$ and $\{\hat{s}_{jt} \leq \frac{\varphi^-}{2J}\}$, so that

$$\begin{aligned}
&\mathbb{P}[|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon] \\
&= \underbrace{\mathbb{P}[|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt} > \frac{\varphi^-}{2J}]\mathbb{P}(\hat{s}_{jt} > \frac{\varphi^-}{2J})}_{H_1} + \underbrace{\mathbb{P}[|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt} \leq \frac{\varphi^-}{2J}]\mathbb{P}(\hat{s}_{jt} \leq \frac{\varphi^-}{2J})}_{H_2}.
\end{aligned}$$

where φ^- is the lower bound from Lemma F.14.

Firstly, for H_1,

$$\mathbb{P}(|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt} > \frac{\varphi^-}{2J})\mathbb{P}(\hat{s}_{jt} > \frac{\varphi^-}{2J}) \leq \mathbb{P}(|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt} > \frac{\varphi^-}{2J}).$$

We therefore would like to bound the value $\mathbb{P}(|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt} > \frac{\varphi^-}{2J})$ and we know

$$|\log \hat{s}_{jt} - \log s_{jt}| = |\log(\frac{\hat{s}_{jt} - s_{jt}}{s_{jt}} + 1)| \leq \max(\hat{s}_{jt}^{-1}, s_{jt}^{-1})|\hat{s}_{jt} - s_{jt}|,$$

where we use the fact that $\frac{x}{1+x} \leq \log(1 + x) \leq x$ for $x > -1$ so $|\log(1 + x)| \leq |\max(x, \frac{x}{1+x})|$ by taking $x = \frac{\hat{s}_{jt}-s_{jt}}{s_{jt}} > -1$. Then we have

$$\begin{aligned}
\mathbb{P}(|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt} > \frac{\varphi^-}{2J}) &\leq \mathbb{P}[\max(\hat{s}_{jt}^{-1}, s_{jt}^{-1})|\hat{s}_{jt} - s_{jt}| > \epsilon \mid \hat{s}_{jt} > \frac{\varphi^-}{2J}] \\
&\leq \mathbb{P}(\frac{2J}{\varphi^-}|\hat{s}_{jt} - s_{jt}| > \epsilon \mid \hat{s}_{jt} > \frac{\varphi^-}{2J}) \\
&= \mathbb{P}(|\hat{s}_{jt} - s_{jt}| > \frac{\varphi^-\epsilon}{2J} \mid \hat{s}_{jt} > \frac{\varphi^-}{2J}).
\end{aligned}$$

We use the law of total probability,

$$\begin{aligned}
\mathbb{P}(|\hat{s}_{jt} - s_{jt}| > \frac{\varphi^-\epsilon}{2J} \mid \hat{s}_{jt} > \frac{\varphi^-}{2J}) &\leq \frac{\mathbb{P}(|\hat{s}_{jt} - s_{jt}| > \frac{\varphi^-\epsilon}{2J})}{\mathbb{P}(\hat{s}_{jt} > \frac{\varphi^-}{2J})} \leq \frac{\mathbb{P}(|\hat{s}_{jt} - s_{jt}| > \frac{\varphi^-\epsilon}{2J})}{1 - e^{-\varphi^{-2}N/8C_1J}} \\
&\lesssim e^{-\varphi^- N\epsilon^2/12J^2 s_{jt}} \leq e^{-\varphi^{-2}N\epsilon^2/6\varphi^+ J}
\end{aligned}$$

where the second inequality is from Lemma *F*.18. With $J = o(N)$, the denominator is bounded by a constant. The first approximated inequality is from Lemma F.19 where we require $\epsilon < 2$ so $\frac{\varphi^-\epsilon}{2J} < \frac{\varphi^-\epsilon}{J} < s_{jt}$.

Secondly, for H_2,

$$\mathbb{P}(|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt} < \frac{\varphi^-}{2J})\mathbb{P}(\hat{s}_{jt} < \frac{\varphi^-}{2J}) \leq \mathbb{P}(\hat{s}_{jt} < \frac{\varphi^-}{2J}) \lesssim e^{-\varphi^{-2}N/8C_1J},$$

where the second inequality is also given by Lemma *F*.18.

Therefore,

$$\mathbb{P}[|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon] \lesssim e^{-\varphi^{-2}N\epsilon^2/6\varphi^+ J} + e^{-\varphi^{-2}N/8\varphi^+ J}. \tag{16}$$

and we can take the minimum of the constants in the exponential to get the desired result. □

With high probability, $\hat{s}_{jt}$ is well defined and $\log \hat{s}_{jt} - \log s_{jt}$ is close to zero. If we drop the not well-defined part of the random variable $\log \hat{s}_{jt} - \log s_{jt}$, it will also be close to zero with high probability.

Lemma F.22. *Given alternative $j \in [J]$ in market $t \in [T]$, assume utilities are bounded such that $\frac{\varphi^-}{J} < s_{jt} < \frac{\varphi^+}{J}$, $J = o(N)$, and $\delta > 0$. For any* $0 < \epsilon < 2$*,*

$$\mathbb{P}[|\mathbb{1}\{\hat{s}_{jt} > \delta\}(\log \hat{s}_{jt} - \log s_{jt})| > \epsilon] < C_1(e^{-C_2N\epsilon^2/J^2} + e^{-C_2N/J^2}),$$

where C_1 and C_2 are positive constants.

Proof. This holds since $|\mathbb{1}\{\hat{s}_{jt} > \delta\}(\log \hat{s}_{jt} - \log s_{jt})| \leq |\log \hat{s}_{jt} - \log s_{jt}|$ and we use Lemma F.21. □

Lemma F.23. *Given an alternative $j \in [J]$ and market $t \in [T]$, assume utilities are bounded such that $\frac{\varphi^-}{J} < s_{jt} < \frac{\varphi^+}{J}$ and $\delta \leq \frac{\varphi^-}{2J}$. For any $\epsilon > 0$,*

$$\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt} \leq \delta\} \log s_{jt}| > \epsilon) < C_1 T(e^{-C_2 N\epsilon^2/J} + e^{-C_2 N/J}).$$

where C_1 and C_2 are positive constants.

Proof. We use a decomposition

$$\begin{aligned}
&\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt} \leq \delta\} \log s_{jt}| > \epsilon) \\
&= \mathbb{P}(|\mathbb{1}\{\hat{s}_{jt} \leq \delta\} \log s_{jt}| > \epsilon \mid \hat{s}_{jt} \geq \frac{\varphi^-}{2J})\mathbb{P}(\hat{s}_{jt} \geq \frac{\varphi^-}{2J}) + \mathbb{P}(|\mathbb{1}\{\hat{s}_{jt} \leq \delta\} \log s_{jt}| > \epsilon \mid \hat{s}_{jt} < \frac{\varphi^-}{2J})\mathbb{P}(\hat{s}_{jt} < \frac{\varphi^-}{2J}) \\
&= \mathbb{P}(|\mathbb{1}\{\hat{s}_{jt} \leq \delta\} \log s_{jt}| > \epsilon \mid \hat{s}_{jt} < \frac{\varphi^-}{2J})\mathbb{P}(\hat{s}_{jt} < \frac{\varphi^-}{2J}) \leq \mathbb{P}(\hat{s}_{jt} < \frac{\varphi^-}{2J}) \leq e^{-\varphi^{-2} N/8C_1 J}
\end{aligned}$$

where the second equality is from $\delta \leq \frac{\varphi^-}{2J}$ and the last inequality is by Lemma F.20 □

Lemma F.24. *Given alternative $j \in [J]$ in market $t \in [T]$, assume utilities are bounded such that $\frac{\varphi^-}{J} < s_{jt} < \frac{\varphi^+}{J}$, $J = o(N)$, and $\delta \leq \frac{\varphi^-}{2J}$. For any $0 < \epsilon < 2$,*

$$\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt} > \delta\} \log \hat{s}_{jt} - \log s_{jt}| > \epsilon) < C_1 T(e^{-C_2 N\epsilon^2/J} + e^{-C_2 N/J}).$$

where C_1 and C_2 are positive constants.

Proof. The inequality comes from the union bound. □

The following lemma suggests that if we sum the errors of the distances between the empirical shares and probabilities over markets, we would like to have the data satisfying $J \log T << N$ to get a small summation.

Lemma F.25. *Given alternative $j \in [J]$ and market $t \in [T]$, assume utilities are bounded such that $\frac{\varphi^-}{J} < s_{jt} < \frac{\varphi^+}{J}$, $J = o(N)$, and $\delta \leq \frac{\varphi^-}{2J}$. For any $0 < \epsilon < 2$,*

$$\mathbb{P}(\frac{1}{T}\sum_t |\mathbb{1}\{\hat{s}_{jt} > \delta\} \log \hat{s}_{jt} - \log s_{jt}| > \epsilon) < C_1 T(e^{-C_2 N\epsilon^2/J} + e^{-C_2 N/J}).$$

where C_1 and C_2 are positive constants.

Proof. we have

$$\mathbb{P}(\sum_t |\mathbb{1}\{\hat{s}_{jt} > \delta\} \log \hat{s}_{jt} - \log s_{jt}| > T\epsilon) \lesssim T(e^{-C_2 N\epsilon^2/J} + e^{-C_2 N/J})$$

by the union bound. □

Then, we show that the distance of some functions of the empirical shares and shares are also close to zero with a high probability. The intuitions are the same as above and provide foundations for the 2SLS estimator.

Lemma F.26. *Given alternatives* $(j, j') \in [J]$ *with bounded utilities such that* $\frac{\varphi^-}{J} < s_{jt} < \frac{\varphi^+}{J}$ *and* $\frac{\varphi^-}{J} < s_{j't} < \frac{\varphi^+}{J}$ *and a uniformly bounded stochastic process* $v_t(w)$ *where* $w \in B$ *and* B *is compact, if* $J = o(N)$ *and* $\varphi \le \frac{\varphi^-}{2J}$*, then for any* $0 < \epsilon < 2$

$$\mathbb{P}\left\{|\sup_{w\in B} \frac{1}{T}\sum_t \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \varphi\}[\left(\log \hat{s}_{jt} - \log \hat{s}_{j't} - v_t(w)\right)^2 - \left(\log s_{jt} - \log s_{j't} - v_t(w)\right)^2]| > \epsilon\right\}$$
$$< C_1 T(e^{-C_2 N\epsilon^2/J\log^2 J} + e^{-C_2 N/J\log^2 J})$$

where C_1 *and* C_2 *are positive constants.*

Proof. The goal is to bound

$$|\sup_{w\in B} \frac{1}{T}\sum_t \mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \varphi\}\{[\log \hat{s}_{jt} - \log \hat{s}_{j't} - v_t(w)]^2 - [\log s_{jt} - \log s_{j't} - v_t(w)]^2\}|.$$

We can instead bound,

$$\sup_{w\in B} \frac{1}{T}\sum_t |\mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \varphi\}\{[\log \hat{s}_{jt} - \log \hat{s}_{j't} - v_t(w)]^2 - [\log s_{jt} - \log s_{j't} - v_t(w)]^2\}|.$$

It is sufficient to check,

$$\sup_{w\in B} \frac{1}{T}\sum_t |\mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \varphi\}\{[\log \hat{s}_{jt} - \log \hat{s}_{j't} - v_t(w)]^2 - [\log s_{jt} - \log s_{j't} - v_t(w)]^2\}|$$
$$= \sup_{w\in B} \frac{1}{T}\sum_t |\mathbb{1}\{\hat{s}_{jt} \wedge \hat{s}_{j't} > \varphi\}[\log \hat{s}_{jt} - \log \hat{s}_{j't} - \log s_{jt} + \log s_{j't}]$$
$$[\log \hat{s}_{jt} - \log \hat{s}_{j't} + \log s_{jt} - \log s_{j't} - 2v(w)]|.$$

Notice

$$
\begin{aligned}
&\mathbb{P}\Big\{\sup_{w\in B}\frac{1}{T}\sum_t |\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\log\hat{s}_{jt}-\log\hat{s}_{j't}-\log s_{jt}+\log s_{j't}]\\
&\qquad\qquad\qquad [\log\hat{s}_{jt}-\log\hat{s}_{j't}+\log s_{jt}-\log s_{j't}-2v(w)]|>\epsilon\Big\}\\
&=\mathbb{P}\Big\{\sup_{w\in B}\sum_t |\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\log\hat{s}_{jt}-\log\hat{s}_{j't}-\log s_{jt}+\log s_{j't}]\\
&\qquad\qquad\qquad [\log\hat{s}_{jt}-\log\hat{s}_{j't}+\log s_{jt}-\log s_{j't}-2v(w)]|>T\epsilon\Big\}\\
&\leq\sup_{w\in B}\mathbb{P}\Big\{\sum_t |\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\log\hat{s}_{jt}-\log\hat{s}_{j't}-\log s_{jt}+\log s_{j't}]\\
&\qquad\qquad\qquad [\log\hat{s}_{jt}-\log\hat{s}_{j't}+\log s_{jt}-\log s_{j't}-2v(w)]|>T\epsilon\Big\}\\
&\leq\sup_{w\in B}\sum_t\mathbb{P}\Big\{|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\log\hat{s}_{jt}-\log\hat{s}_{j't}-\log s_{jt}+\log s_{j't}]\\
&\qquad\qquad\qquad [\log\hat{s}_{jt}-\log\hat{s}_{j't}+\log s_{jt}-\log s_{j't}-2v(w)]|>\epsilon\Big\}
\end{aligned}
$$

where the first and second inequalities are from the union bound. Then, we have

$$
\begin{aligned}
&\sup_{w\in B}\sum_t\mathbb{P}\Big\{|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\log\hat{s}_{jt}-\log\hat{s}_{j't}-\log s_{jt}+\log s_{j't}]\\
&\qquad\qquad\qquad [\log\hat{s}_{jt}-\log\hat{s}_{j't}+\log s_{jt}-\log s_{j't}-2v(w)]|>\epsilon\Big\}\\
&=\sup_{w\in B}\sum_t\mathbb{P}\Big\{|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\log\hat{s}_{jt}-\log\hat{s}_{j't}-\log s_{jt}+\log s_{j't}]\\
&\qquad [\log\hat{s}_{jt}-\log\hat{s}_{j't}+\log s_{jt}-\log s_{j't}-2v(w)]|>\epsilon\mid\hat{s}_{jt}\wedge\hat{s}_{j't}>\frac{2c_0}{J}\Big\}\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}>\frac{2c_0}{J})\\
&+\sup_{w\in B}\sum_t\mathbb{P}\Big\{|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\log\hat{s}_{jt}-\log\hat{s}_{j't}-\log s_{jt}+\log s_{j't}]\\
&\qquad [\log\hat{s}_{jt}-\log\hat{s}_{j't}+\log s_{jt}-\log s_{j't}-2v(w)]|>\epsilon\mid\hat{s}_{jt}\wedge\hat{s}_{j't}<\frac{2c_0}{J}\Big\}\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}<\frac{2c_0}{J}).
\end{aligned}
$$

Therefore, to bound the sum, we need to concentrate on

$$
\begin{aligned}
H_1 = &\sup_{w\in B}\sum_t \mathbb{P}\Big\{|\mathbb{1}\{\hat{s}_{jt}\wedge \hat{s}_{j't} > \varphi\}[\log \hat{s}_{jt} - \log \hat{s}_{j't} - \log s_{jt} + \log s_{j't}] \\
&\quad [\log \hat{s}_{jt} - \log \hat{s}_{j't} + \log s_{jt} - \log s_{j't} - 2v(w)]| > \epsilon \mid \hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J}\Big\}\mathbb{P}(\hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J}) \\
H_2 = &\sup_{w\in B}\sum_t \mathbb{P}\Big\{|\mathbb{1}\{\hat{s}_{jt}\wedge \hat{s}_{j't} > \varphi\}[\log \hat{s}_{jt} - \log \hat{s}_{j't} - \log s_{jt} + \log s_{j't}] \\
&\quad [\log \hat{s}_{jt} - \log \hat{s}_{j't} + \log s_{jt} - \log s_{j't} - 2v(w)]| > \epsilon \mid \hat{s}_{jt}\wedge \hat{s}_{j't} < \frac{2c_0}{J}\Big\}\mathbb{P}(\hat{s}_{jt}\wedge \hat{s}_{j't} < \frac{2c_0}{J}).
\end{aligned}
$$

We focus on H_1 first

$$
\begin{aligned}
H_1 \le\ & T\sup_{w\in B}\mathbb{P}\Big\{|\mathbb{1}\{\hat{s}_{jt}\wedge \hat{s}_{j't} > \varphi\}|\log \hat{s}_{jt} - \log \hat{s}_{j't} - \log s_{jt} + \log s_{j't}| \\
&\quad (|\log \hat{s}_{jt}| + |\log \hat{s}_{j't}| + |\log s_{jt}| + |\log s_{j't}| + |2v(w)|)| > \epsilon \mid \hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J}\Big\}\mathbb{P}(\hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J}) \\
\le\ & T\sup_{w\in B}\mathbb{P}\Big\{|\log \hat{s}_{jt} - \log \hat{s}_{j't} - \log s_{jt} + \log s_{j't}| \\
&\quad (|\log\frac{2c_0}{J}| + |\log\frac{2c_0}{J}| + |\log\frac{c_0}{J}| + |\log\frac{c_0}{J}| + 2c_2) > \epsilon \mid \hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J}\Big\} \\
\le\ & T\mathbb{P}(C_3\log J|\log \hat{s}_{jt} - \log \hat{s}_{j't} - \log s_{jt} + \log s_{j't}| > \epsilon \mid \hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J}) \\
\lesssim\ & 2T\mathbb{P}(C_3\log J|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon \mid \hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J}) \\
\le\ & \frac{2T\mathbb{P}(C_3\log J|\log \hat{s}_{jt} - \log s_{jt}| > \epsilon)}{\mathbb{P}(\hat{s}_{jt}\wedge \hat{s}_{j't} > \frac{2c_0}{J})} \le \frac{2T\mathbb{P}(|\log \hat{s}_{jt} - \log s_{jt}| > \frac{\epsilon}{C_3\log J})}{1 - 2\mathbb{P}(\hat{s}_{jt} > \frac{2c_0}{J})} \lesssim T(e^{-C_2N\epsilon^2/J\log^2 J} + e^{-C_2N/J})
\end{aligned}
$$

where the fourth inequality is from the law of total probability. The last inequality relies on Lemma F.21 and $J = o(n)$. Note that there is no guarantee that $\frac{\epsilon}{C_3\log J} < 2$ for any positive constant C_3, but we can take C_3 large. $c_2 > 0$ is the upper bound of $|v_t(w)|$ and $C_3 > 0$ is a constant. $C_1 > 0$ and $C_2 > 0$ are constants which do not depend on j and t. Then we discuss H_2:

$$
H_2 \le T\mathbb{P}(\hat{s}_{jt}\wedge \hat{s}_{j't} < \frac{2\varphi^-}{J}) \lesssim Te^{-\tilde{C}^2N/J},
$$

where the approximated inequality is from Lemma F.20 and $\tilde{C}_2 > 0$ is a constant.

Therefore, by the union bound, we have

$$
\begin{aligned}
&\mathbb{P}\left\{|\sup_{w\in B}\frac{1}{T}\sum_t \mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}[\left(\log\hat{s}_{jt}-\log\hat{s}_{j't}-v_t(w)\right)^2-\left(\log s_{jt}-\log s_{j't}-v_t(w)\right)^2]|>\epsilon\right\}\\
&\leq C_1T(e^{-C_2N\epsilon^2/J\log^2 J}+e^{-C_2N/J}).
\end{aligned}
$$

where C_1 and C_2 are positive constants. □

Lemma F.27. *Given alternatives $(j,j')\in[J]$ with bounded utilities such that $\frac{\varphi^-}{J}<s_{jt}<\frac{\varphi^+}{J}$ and $\frac{\varphi^-}{J}<s_{j't}<\frac{\varphi^+}{J}$ and a uniformly bounded stochastic process $v_t(w)$ where $w\in B$ and B is compact, if $\varphi\leq\frac{\varphi^-}{2J}$, then for any $0<\epsilon<2$*

$$
\mathbb{P}(\sup_{w\in B}\frac{1}{T}\sum_t|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}<\varphi\}\{[\log s_{jt}-\log s_{j't}-v_t(w)]^2\}|>\epsilon)<C_1e^{-C_2N/J}
$$

with a positive constant C_1.

Proof. By the law of total probability,

$$
\begin{aligned}
&\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}<\varphi\}\{[\log s_{jt}-\log s_{j't}-v_t(w)]^2\}|)\\
&=\underbrace{\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}<\varphi\}\{[\log s_{jt}-\log s_{j't}-v_t(w)]^2\}|\mid\hat{s}_{jt}\wedge\hat{s}_{j't}>\frac{\varphi^-}{2J})\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}>\frac{\varphi^-}{2J})}_{H_1}\\
&+\underbrace{\mathbb{P}(|\mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}<\varphi\}\{\log s_{jt}-\log s_{j't}-v_t(w)]^2\}|\mid\hat{s}_{jt}\wedge\hat{s}_{j't}<\frac{\varphi^-}{2J})\mathbb{P}(\hat{s}_{jt}\wedge\hat{s}_{j't}<\frac{\varphi^-}{2J})}_{H_2}.
\end{aligned}
$$

Since $\varphi<\frac{\varphi^-}{2J}$, we have $H_1=0$. And

$$
H_2\leq\mathbb{P}(\hat{s}_{jt}<\frac{\varphi^-}{2J})\leq e^{-\varphi^{-2}N/8c_1J}
$$

Then, we use the union bound inequality to derive the result. □

Lemma F.28. *Given alternatives j and j' with bounded utilities and a uniformly bounded stochastic process $v_t(w)$ where $w\in B$. If $J=o(N)$ and $\varphi\leq\frac{c_0}{2J}$, then for any $0<\epsilon<2$*

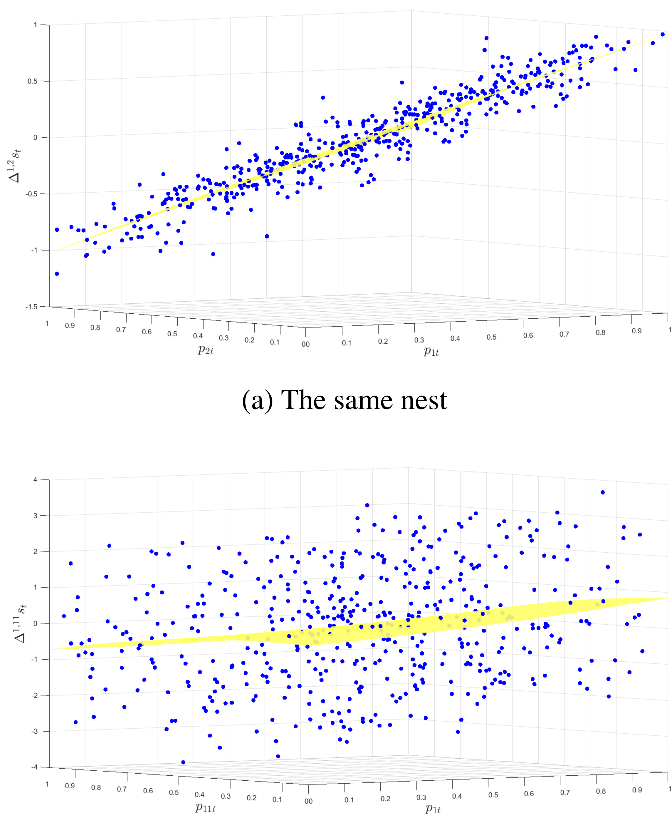
$$
\begin{aligned}
&\mathbb{P}\left\{|\sup_{w\in B}\frac{1}{T}\sum_t \mathbb{1}\{\hat{s}_{jt}\wedge\hat{s}_{j't}>\varphi\}\left[\log\hat{s}_{jt}-\log\hat{s}_{j't}-v_t(w)\right]^2-\left[\log s_{jt}-\log s_{j't}-v_t(w)\right]^2|>\epsilon\right\}\\
&<C_1T(e^{-C_2N\epsilon^2/J\log^2 J}+e^{-C_2N/J})
\end{aligned}
$$

where C_1 and C_2 are positive constants.

Proof. This is a direct application of the union bound. □

G Supplementary figures

Figure 6: Polynomial surfaces and points for same vs. different nest pairs (Nested Logit)



(a) The same nest

(b) Different nests

H Connect FMML to Topic Model

Example 3 (Topic Model) In natural language processing, the most common interpretable models are topic models (Ash and Hansen, 2023). We observe documents $t \in [T]$ based on a dictionary of $j \in [J]$ terms. There are latent topics $k \in [K]$ for generating the documents. We assume that the probability of a term j appearing in a given entry of document t, s_{jt}, is given by

$$s_{jt} = \sum_{k=1}^{K} \mathbb{P}_t(\text{Topic } k)\mathbb{P}(\text{Term } j \mid \text{Topic } k) = \sum_{k=1}^{K} \pi_{kt} A_{kj},$$

or equivalently,

$$P = A^\top W,$$

where $P \in \mathbb{R}^{J\times T}$, $A \in \mathbb{R}^{K\times J}$, and $W \in \mathbb{R}^{K\times T}$ with $P_{jt} = s_{jt}$ and $W_{kt} = \pi_{kt}$. In Appendix I.3 we show that for any set of $\{A_{kj}\}$, there exists $\{u_{kj}\}$ such that $A_{kj} = \frac{\mathbb{1}_{J_k}(j)e^{u_{kj}}}{\sum_{i\in J_k} e^{u_{ik}}}$ for $j \in [J]$ and $k \in [K]$. Specifically, their models ignore covariate x_{jt} while our models allow for variations in the probabilities of terms conditional on topics by the logit models. Such an extension is meaningful. In economic applications of policy communications, the words used in the documents rely on the economic conditions and political parties. Violating the basic assumptions of the topic models can bring about difficulty in interpretations of the results from misspecification (Ke et al., 2019).

I Identification without unobserved characteristics

I.1 Extra notation

For a generic matrix $Q \in \mathbb{R}^{d\times m}$, we let $Q_{i\cdot}$ and $Q_{\cdot j}$ be the ith row and jth column of Q. Finally, write the $d \times d$ diagonal matrix

$$D_Q := \text{diag}(\|Q_{1\cdot}\|_1, \ldots, \|Q_{d\cdot}\|_1)$$

and let $(D_Q)_{ii}$ denote the i-*th* diagonal element. Finally, we denote $Q^{\text{row}} := D_Q^{-1} Q$ as the row normalization of matrix Q.

I.2 Non-identification with unobserved characteristics

Here, we present an illustrated example demonstrating that identical observed probabilities can be depicted through two different types of mixtures. The matrix entries denote the probabilities of products occurring within these mixtures, with four products represented in the rows and three mixture types. With unobserved characteristics, the matrix values in A are arbitrary.

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & & \\ & \frac{1}{2} & \\ & & \frac{1}{2} \end{bmatrix} \text{ and } W = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}, \text{ we have } AW = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \end{bmatrix}.$$

For an alternative representation,

$$\tilde{A} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & & \\ & \frac{2}{3} & \\ & & \frac{1}{3} \end{bmatrix} \text{ and } \tilde{W} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{2} \end{bmatrix}, \text{ we also have } \tilde{A}\tilde{W} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \end{bmatrix}$$

This example underscores that, even with knowledge about the product-mixture associations, determining the probabilities of mixtures and utilities requires additional assumptions.

I.3 Connection to topic model

We show that for any given set of $\{A_{kj}\}$, there exists a set of $\{u_{kj}\}$ such that for all $j \in [J]$ and $k \in K$, the following holds:

$$A_{kj} = \frac{\mathbb{1}_{J_k}(j)\, e^{u_{kj}}}{\sum_{i \in J_k} e^{u_{ki}}}.$$

Without loss of generality, we fix k for the analysis. The set J_k is specified as $J_k = \{j \in J \mid A_{kj} > 0\}$. Also, for any j where $A_{kj} > 0$, let $u_{kj} = \log A_{kj}$. This specification supports the argument that $A_{kj} = \frac{\mathbb{1}_{J_k}(j)\, e^{u_{kj}}}{\sum_{i \in J_k} e^{u_{ki}}}$ because $\sum_{i \in J_k} A_{ki} = 1$.

I.4 Identification assumption: Seperability

Definition 6. *A column stochastic, rank K matrix $A \in \mathbb{R}^{K \times J}$ is said to be separable if there exists a row permutation matrix Π such that*

$$\Pi A^\top = \begin{bmatrix} D \\ M \end{bmatrix}, \tag{17}$$

where $D \in \mathbb{R}^{K \times K}$ is a diagonal nonnegative matrix.

It will be convenient to have an explicit definition of what it means to say that P *admits a nonnegative separable matrix factorization*:

Definition 7. *A column stochastic matrix $P \in \mathbb{R}^{J \times T}$ with nonnegative rank K is said to have a rank K separable (or anchor-word) factorization if P can be written as $P = A^\top W$, where $A \in \mathbb{R}^{K \times J}$ is some matrix that satisfies Definition 6, and W is a $K \times D$ column stochastic matrix.*

I.5 Identification without unobserved characteristics

Definition 8. *A* loner *of a row-normalized matrix is a row r which is not a convex combination of at least two rows, r', r'', with $r \neq r'$ and $r \neq r''$.*

Assume $p_{jt} = p_j$ for $t \in [T]$ in (3) that there is no variation of the prices in the markets. We denote $A_{kj} = s_j^*(p|C_k)$ and $W_{kt} = \pi_{kt}$. If both A and W have full rank K and A satisfies the separability assumption, then model (3) is identified up to permutations.

Proof. We let $s = A^\top W$ rely on the row normalization of share matrix s. Let $A^{*\top} = D_P^{-1} A^\top D_W$ then $P^{\text{row}} = A^{*\top} W^{\text{row}}$. And we rely on the loners for the identification. W is of full rank, and D_W is of full rank by the nonnegativity. Therefore, W^{row} is of full rank as well, and the rows in the W^{row} are loners. Then, we want to show that the loners in P^{row} are rows in W^{row}.

Firstly, we show the rows in the P^{row} are convex combinations of the rows in W^{row}. This is because $A^* \mathbb{1}_K = A^* W^{\text{row}} \mathbb{1}_T = P^{\text{row}} \mathbb{1}_T = \mathbb{1}_D$. We then show that if a row in A represents an anchor word, the corresponding row in $A^{*\top}$ is a unit vector on a standard basis. Right multiplication of D_P^{-1} multiplies the rows by fixed numbers, and left multiplication of D_W multiplies the columns by fixed numbers. Neither of the operations changes the sparsity pattern of the matrix.

With the rows in P^{row} are linear combinations of row in W^{row}, we prove that the all rows in W^{row} can be found by loners in P^{row}. We prove this by contradiction. As we discussed, due to that each anchor word exists for each topic and therefore the sparsity pattern of $A^{*\top}$, all rows of W^{row} appear in P^{row}. If W_1^{row} is a strictly positive linear combination of $\{P_{m_1}^{\text{row}}, \cdots, P_{m_l}^{\text{row}}\}$ and $P_{m_1}^{\text{row}} \neq W_1^{\text{row}}$. That is $W_1^{\text{row}} = \sum_{i=1}^{l} a_i P_{m_i}^{\text{row}}$ and $a_i > 0$. Since $\{P_{m_1}^{\text{row}}, \cdots, P_{m_l}^{\text{row}}\}$ are linear combinations of $\{W_1^{\text{row}}, \cdots, W_k^{\text{row}}\}$, $P_{m_i}^{\text{row}} = \sum_{j=1}^{k} a_{ij} W_j^{\text{row}}$ for $a_{ij} \geq 0$ and there exists $j' \geq 2$ s.t. $a_{1j'} > 0$. Then

$$W_1^{\text{row}} = \sum_{i=1}^{l} a_i \sum_{j=1}^{k} a_{ij} W_j^{\text{row}} = \sum_{j=1}^{k} \sum_{i=1}^{l} a_i a_{ij} W_{j\bullet}^{\text{row}}.$$

The weight of $W_{j'}^{\text{row}}$ is larger than 0 so W_1^{row} is not a loner. This is a contradiction.

For the rows other than W^{row} in P^{row}, they are all linear combinations of rows in $\tilde{W}$. Therefore the loners in P^{row} are rows in W^{row}. We can recover W^{row} by checking the loners in P^{row}. We can derive A^* by $A^* = P^{\text{row}} W^{\text{row}+}$. Then A can be calculated by column normalizing $D_p A^*$. □

I.6 Separability with unobserved characteristics

For $P_1 = A_1^\top W_1$ with dimension $J \times T_1$ and $P_2 = A_2^\top W_2$ with dimension $J \times T_2$, we assume all A_1, A_2, W_1, and W_2 have full rank K. The anchor pattern of A_1 and A_2 are the same but $A_1 \neq A_2$. Then there does not exist NMF under the separability assumption for $P = [P_1, P_2]$

Proof. We prove this by contradiction. We will show that $A_1 = A_2$ if NMF for P exists under the separability assumption. Assume there exists $P = A^\top W$ and A satisfies the anchor word factorization. We have $A^\top \tilde{W}_1 = A_1 W_1$ and $A^\top \tilde{W}_2 = A_2 W_2$ where $\tilde{W}_1$ is the first T_1 columns of W and $\tilde{W}_2$ is the last T_2 columns of W. Since A_1 and A_2 satisfy the anchor word assumption, by the uniqueness of the anchor word factorization for P_1 and P_2, we have $A_1 = A = A_2$, which is a contradiction. $\square$