Regularized Demand Inversion with Sampling Zeros in Market Share*

Ertian Chen† Lars Nesheim‡

November 6, 2026

Abstract

Demand estimation often requires inversion of a demand system to recover unobserved product-market demand shocks. Zero observed market shares make standard inversion undefined. We propose regularized demand inversion, which places a quadratic penalty on demand shocks in the joint multinomial log-likelihood. The regularized inverse exists even with sampling zeros and is computationally fast. We establish finite-sample recovery bounds that guide the choice of the regularization parameter. We show that the resulting GMM estimator is first-order asymptotically equivalent to the feasible sample-share GMM estimator defined when observed shares are strictly positive. Monte Carlo simulations show that the estimator performs well even with extremely many sampling zeros. We apply the method to the London beer market and find that accounting for zeros has a large impact on the estimated price elasticities.

Keywords: Demand estimation; sampling zeros; differentiated products.

JEL codes: C13, C25, C51, L13.

*The use of Worldpanel by Numerator data does not imply the endorsement of Worldpanel by Numerator in relation to the interpretation or analysis of the data. All errors are our own.

†School of Economics & Paula and Gregory Chow Institute for Studies in Economics, Xiamen University and IFS. Email: ertianchen@xmu.edu.cn.

‡University College London and IFS. Email: l.nesheim@ucl.ac.uk.

1 Introduction

Zero market shares are common in demand estimation with granular data. Data from consumer scanner panels, online retail, and automobile registrations usually contain many products that are available but not purchased by any sampled consumer (see, e.g., Quan and Williams, 2018; D'Haultfœuille et al., 2019; Li, 2023). Standard demand estimation following Berry (1994) and Berry et al. (1995, hereafter BLP) recovers demand shocks by matching model-implied shares to observed shares and then instrumenting the recovered shocks to estimate structural parameters. However, the presence of zeros makes the usual inversion undefined, which leaves an important gap in the BLP framework. This paper studies sampling zeros, which arise when no purchase of a product is recorded in a finite sample even though the product has a positive true market share. Such zeros can remain common even in large samples when purchases are spread across many products.

We propose regularized demand inversion (RDI). RDI builds on a maximum-likelihood interpretation of demand inversion. When every observed share, including the outside share, is strictly positive, the shocks recovered from observed shares are the maximum-likelihood estimates under the multinomial sampling model, holding the structural parameters fixed. When an observed share is zero, the likelihood has no finite maximizer. RDI subtracts a quadratic penalty on these shocks from the log-likelihood and recovers the shocks that maximize the penalized likelihood. The regularization ensures that a finite solution exists for every observed share vector, including those with zeros. The regularization acts on the demand shocks rather than altering observed shares.[1] This distinction is economically useful because the demand shocks are primitives of the demand model, whereas market shares are equilibrium outcomes determined jointly by product characteristics, competitors, and the choice set.

We establish finite- and large-sample properties of the regularized inverse. We first derive a bound on its mean squared error relative to the shocks implied by the true shares, conditional on those shares and the number of sampled consumers. The bound is uniform over product characteristics and structural parameters and applies to every regularized inverse, including cases with multiple solutions. Minimizing its leading terms gives a rate-balancing scale of the regularization parameter that depends only on the numbers of sampled consumers and products. Holding the number of products fixed, a smaller sample raises both this choice and the expected fraction of products with zero observed shares. In finite samples, the regularized inverse is unique for every observed share vector when market shares

[1] See Section 1.1 for a discussion of alternative approaches to sampling zeros.

are log-concave in the demand shocks, as in simple logit and in random-coefficients logit and nested logit with normal or bounded log-concave tastes. With probability approaching one, the regularized inverse is unique and converges uniformly to the shocks implied by the true shares.

Our estimator uses the regularized inverse in the standard instrumental-variable moments to estimate structural parameters. We establish consistency and first-order asymptotic equivalence to the feasible BLP estimator defined when observed shares are strictly positive, thereby showing our estimator is asymptotically normal. The asymptotic results hold both for a fixed regularization scale and for the rate-balancing scale.

Computationally, researchers only need to add a quadratic penalty to the log-likelihood and then use standard optimization routines to recover the regularized inverse. Therefore, the estimation is straightforward to implement.

We apply our method to Monte Carlo simulations based on Gandhi et al. (2023). We consider exogenous and endogenous product characteristics and four demand-shock distributions: Gaussian, Student-t, contaminated normal, and bimodal. RDI performs well in all designs, even with extremely many sampling zeros. Stronger regularization is especially helpful in the endogenous designs with many zeros. Increasing the number of markets improves parameter estimates but has little effect on the accuracy of individual demand shocks.

We apply RDI to the London beer market using Worldpanel by Numerator's Take Home Purchase Panel during 2015–2019. We define a market as a calendar week. On average, each market has 67% of products with zero observed shares. We estimate a random-coefficients nested logit model similar to that of Miller and Weinberg (2017). RDI gives a mean own-price elasticity of -3.32, compared with -1.20 after dropping zero cells, -1.29 with Laplace smoothing, and -0.44 with Gandhi et al. (2023). Therefore, our method performs well in practice and has a large impact on estimated price elasticities. In the counterfactual analysis, we study the effect of applying the 2026 alcohol-duty schedule to the London beer market in each year of 2015–2019.

1.1 Related Literature

There are two strands of related literature in econometrics. The first studies sampling zeros and the second studies true zeros. The literature on sampling zeros usually works on the observed shares. D'Haultfœuille et al. (2019) add a small constant to observed shares, Li (2023) imposes a prior on the true shares and uses similar markets to estimate the prior parameters, Lieber and Wiemann (2022) consider probabilistic bounds on the true shares,

Gandhi et al. (2023) impose a lower bound on the true shares and construct bounds on the shocks to use moment inequalities, and Hortaçsu et al. (2023) impose a distributional assumption on the demand shocks and pricing equation shocks. Adam et al. (2024) correct for selection bias from excluding products with zero observed shares. Quan and Williams (2018) treat demand shocks in local markets as random effects and invert aggregate demand to recover national-level demand shocks. Other approaches restrict unobserved demand through sparsity (Lu and Shimizu, 2025), control functions (Stourm and Stourm, 2025), or additive fixed effects (Dubois et al., 2020; O'Connell and Smith, 2024). Allen (2026) studies identification when market size is bounded. A recent paper by Hu (2026) treats infinite values from demand inversion with zeros as outliers and uses the least absolute deviation estimator to handle the outliers. Our method is different from these approaches because we regularize the demand shocks rather than the observed shares.

True zeros arise when the demand system allows for zero model-implied shares. Examples include changing the distribution of utility shocks (Berry and Pakes, 2007; de Lima, 2024) and using consideration sets (Dubé et al., 2021).

Our RDI approach is related to the literature on regularized maximum likelihood estimation. Albert and Anderson (1984); King and Zeng (2001) study existence and rare-event bias in logistic regression. Firth (1993); Heinze and Schemper (2002); Kosmidis and Firth (2021) study bias reduction in maximum likelihood estimation by penalizing the log-likelihood. The regularization can be interpreted as a prior on the parameters, which is also used in the panel data literature (Arellano and Bonhomme, 2009; Yan et al., 2026; Beyhum et al., 2026).

Our method also contributes to the literature on BLP-type demand estimation. Berry et al. (2004) study asymptotics with a large number of products and Armstrong (2016) studies BLP-instrument strength with a large number of products and markets. Freyberger (2015) establishes asymptotic theory for demand estimation with many markets, allowing for both sampling error in observed shares and simulation error in model-implied shares. We complement this literature by establishing conditions under which regularized demand inversion preserves the first-order asymptotic distribution of feasible BLP estimation. Therefore, our method can be used in conjunction with the asymptotic theory developed in those papers.

Outline: Section 2 presents RDI and its single-market properties. Section 3 gives the GMM results. Sections 4 and 5 present the simulations and beer application. Section 6 concludes. Proofs and model-specific results appear in the appendices.

Notation: We use $\Delta^J = \left\{ s \in \mathbb{R}_+^{J+1} : \sum_{j=0}^{J} s_j = 1 \right\}$ to denote the J-dimensional unit simplex and use $\text{int}(\Delta^J)$ to denote its relative interior. For a vector $v \in \mathbb{R}^K$, we write $\|v\|_2$ for

its Euclidean norm and $\|v\|_\infty$ for its maximum absolute value. For a matrix $A \in \mathbb{R}^{K\times L}$, we write $\|A\|_2$ for its spectral norm and $\|A\|_\infty$ for its induced maximum-row-sum norm. We write $[A]_{j\cdot}$ for its jth row. For a scalar-valued function f, we write $\partial_\xi f$ for its gradient and $\partial^2_{\xi\xi'} f$ for its Hessian. For a vector-valued function f, $\partial_\xi f$ denotes its Jacobian.

2 Framework

This section introduces regularized demand inversion for canonical Berry et al. (1995) discrete-choice demand models. We establish finite- and large-sample properties of the regularized inverse.

2.1 Regularized Demand Inversion

We consider a market with $J \geq 2$ inside products indexed by $j \in \mathcal{J}_+ := \{1, \ldots, J\}$ and an outside option indexed by $j = 0$. The choice set is $\mathcal{J} = \mathcal{J}_+ \cup \{0\}$. Each inside product j is characterized by a vector $x_j \in \mathcal{X} \subseteq \mathbb{R}^K$ of observed product characteristics and an unobserved product characteristic $\xi_j \in \mathbb{R}$. Let the vector of all observed characteristics be $x = (x_1, \ldots, x_J) \in \mathcal{X}^J \subseteq \mathbb{R}^{KJ}$ and the vector of unobserved characteristics be $\xi = (\xi_1, \ldots, \xi_J)$. There is a vector of structural parameters $\theta \in \Theta \subset \mathbb{R}^d$. We first impose conditions on the market share function.

Assumption 1 (Share-function regularity)**.** *Population market shares are given by the function $\sigma : \mathcal{X}^J \times \mathbb{R}^J \times \Theta \to \text{int}(\Delta^J)$. The function σ is three times continuously differentiable in ξ and twice continuously differentiable jointly in (x, ξ, θ).*

This assumption formalizes the idea that we restrict our analysis to settings with strictly positive population market shares and imposes higher-order differentiability conditions.

We next impose conditions adapted from Berry et al. (2013). For $k \in \mathcal{J}_+$ and $j \in \mathcal{J} \setminus \{k\}$, say that product k is a strict substitute for product j at (x, ξ, θ) when $\frac{\partial \sigma_j(x,\xi;\theta)}{\partial \xi_k} < 0$.

Assumption 2 (Connected substitutes)**.** *For every (x, θ) and $\xi \in \mathbb{R}^J$:*

(i) $\frac{\partial \sigma_j(x,\xi;\theta)}{\partial \xi_k} \leq 0$ *for every $k \in \mathcal{J}_+$ and $j \in \mathcal{J} \setminus \{k\}$;*

(ii) for every nonempty $K \subseteq \mathcal{J}_+$, there are $k \in K$ and $j \in \mathcal{J} \setminus K$ such that $\frac{\partial \sigma_j(x,\xi;\theta)}{\partial \xi_k} < 0$.

Part (i) is weak substitution. Part (ii) is the differentiable connected strict-substitution condition (Assumption 3* of Berry et al. (2013)). Given differentiability and weak substitution, it is equivalent to nonsingularity of the inside-share Jacobian and is only slightly stronger than their qualitative connected-substitution condition.

Assumption 3 (Share-function surjectivity)**.** *For every* $(x, \theta) \in \mathcal{X}^J \times \Theta$,

$$\left\{\sigma(x, \xi; \theta) : \xi \in \mathbb{R}^J\right\} = \operatorname{int}(\Delta^J).$$

Under Assumption 2, Theorems 1–2 and Corollary 1 of Berry et al. (2013) imply that the inside-share map is injective and its Jacobian is nonsingular. Together with Assumption 3, this gives a unique inverse for every $(x, \theta, s) \in \mathcal{X}^J \times \Theta \times \operatorname{int}(\Delta^J)$. We denote it by $\xi^{\mathrm{BLP}}(x, \theta, s) \in \mathbb{R}^J$, with jth component $\xi_j^{\mathrm{BLP}}(x, \theta, s)$, so that

$$\sigma(x, \xi^{\mathrm{BLP}}(x, \theta, s); \theta) = s.$$

We start by noting that demand inversion can be interpreted as a maximum-likelihood problem. The log-likelihood of a market share vector s is

$$L_s(x, \xi; \theta) := \sum_{j \in \mathcal{J}} s_j \log \sigma_j(x, \xi; \theta). \tag{1}$$

Moreover, $\xi^{\mathrm{BLP}}(x, \theta, s)$ maximizes (1) over $\xi \in \mathbb{R}^J$ for every $s \in \operatorname{int}(\Delta^J)$.

For the rest of this section, fix x and suppress it where doing so causes no confusion. We observe N independent consumer choices, each with choice probabilities given by the true market shares $s^0 \in \operatorname{int}(\Delta^J)$. Let $Q = (Q_0, \ldots, Q_J)$ denote the vector of counts of consumers choosing each product, with $\sum_{j \in \mathcal{J}} Q_j = N$. It follows that

$$Q \sim \text{Multinomial}\big(N; s^0\big)\,.$$

The observed market shares are $\hat{s}_j = \frac{Q_j}{N}$ for $j \in \mathcal{J}$.

When some observed market shares are zero, the log-likelihood does not have a finite maximizer. To resolve this problem, we introduce a regularization term that penalizes large values of the unobserved product characteristics ξ. Let $\lambda_{N,J} > 0$ be a deterministic regularization parameter that may depend on N and J. Define the regularized criterion by

$$L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}) := L_s(x, \xi; \theta) - \frac{\lambda_{N,J}}{2} \|\xi\|_2^2. \tag{2}$$

The regularization term[2] "shrinks" ξ toward zero and as a result, as we show below, the regularized criterion has at least one finite maximizer. Larger values of $\lambda_{N,J}$ impose stronger regularization. When some shares are zero, there can be multiple maximizers but we provide bounds on the finite-sample distance between any maximizer and the BLP inverse evaluated at the true shares. Furthermore, we show that, asymptotically, the regularized inverse converges to the BLP inverse. For this asymptotic result we let $\lambda_{N,J} \to 0$.

Because the argmax of the regularized criterion may not be a singleton when some sample counts are zero, we define the regularized demand inverse as follows:

Definition 1 (Regularized demand inverse)**.** *The regularized demand inverse is a fixed measurable selection $\xi_N^{\mathrm{R}}(x, \theta, s) \in \mathcal{A}_N(x, \theta, s)$, where*

$$\mathcal{A}_N(x, \theta, s) = \operatorname*{argmax}_{\xi \in \mathbb{R}^J} L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}).$$

That is, ξ_N^{R} is a Borel measurable map from $\mathcal{X}^J \times \Theta \times \Delta^J$ to $\mathbb{R}^J$ such that, for every (x, θ, s), $\xi_N^{\mathrm{R}}(x, \theta, s) \in \mathcal{A}_N(x, \theta, s)$. Write $\xi_{j,N}^{\mathrm{R}}(x, \theta, s)$ for its jth component.

2.2 Pointwise Existence and Uniqueness

Proposition 1 (Existence and positive-share uniqueness)**.** *Suppose Assumptions 1–3 hold.*

(i) For every $s \in \Delta^J$ and $\lambda_{N,J} > 0$, $\mathcal{A}_N(x, \theta, s)$ is nonempty and compact.

(ii) Fix (x, θ) and $s \in \operatorname{int}(\Delta^J)$. There is $\bar{\lambda}(x, \theta, s) > 0$ such that the regularized inverse is unique whenever $\lambda_{N,J} < \bar{\lambda}(x, \theta, s)$. If $\theta \in \operatorname{int}(\Theta)$, the regularized inverse is locally continuously differentiable in θ.

Proposition 1 is pointwise in (x, θ, s). Because the objective is jointly continuous and its quadratic penalty makes its upper level sets compact, the measurable maximum theorem and a standard measurable-selection result ensure that the selection in Definition 1 exists. The results below in fact hold for every element of the argmax correspondence, so they do not depend on the selection rule.

[2]For $s = \hat{s}$, multiplying (2) by N gives the regularized count log-likelihood $\sum_{j \in \mathcal{J}} Q_j \log \sigma_j(x, \xi; \theta) - \frac{N\lambda_{N,J}}{2}\|\xi\|_2^2$. The regularized inverse is therefore the posterior mode of ξ under independent Gaussian priors $\xi_j \sim \mathcal{N}\big(0, \frac{1}{N\lambda_{N,J}}\big)$. In Section 4, we set $\lambda_{N,J} = \frac{1}{N\tau^2}$, so τ^2 is the prior variance, and we vary both τ and the true distribution of ξ to check the robustness of the RDI estimator.

2.3 Recovery of the Demand Shocks

For constants $0 < c < C < \infty$ and $c_0 > 0$, define

$$\mathcal{S}_J := \left\{ s \in \Delta^J : \frac{c}{J} \leq s_j \leq \frac{C}{J} \ (j \in \mathcal{J}_+), \quad s_0 \geq c_0 \right\}$$

and

$$\mathcal{S}_J^+ := \left\{ s \in \Delta^J : \frac{c}{2J} \leq s_j \leq \frac{2C}{J} \ (j \in \mathcal{J}_+), \quad s_0 \geq \frac{c_0}{2} \right\}.$$

Theorem 1 and Corollary 1 below do not require compactness. We use the following condition to verify Assumption 5 for the logit models in Appendix B and for the GMM results in Section 3.

Assumption 4. *The sets $\mathcal{X}$ and Θ are compact.*

Assumption 5 (Uniform inverse and likelihood regularity)**.** *Write $H_s(x, \xi; \theta) := -\partial^2_{\xi\xi'} L_s(x, \xi; \theta)$. There are constants $C_\star \geq 1$ and $0 < r_0 < \frac{1}{4C_\star^2}$ such that the following conditions hold uniformly over $x \in \mathcal{X}^J$, $s \in \mathcal{S}_J^+$, and $\theta \in \Theta$.*

(i) $\|\xi^{\mathrm{BLP}}(x, \theta, s)\|_\infty \leq C_\star \log J$;

(ii) $\|H_s(x, \xi^{\mathrm{BLP}}(x, \theta, s); \theta)^{-1}\|_\infty \leq C_\star J$ *and, whenever ξ, ξ' are within sup-norm distance r_0 of this inverse,*

$$\|H_s(x, \xi; \theta) - H_s(x, \xi'; \theta)\|_\infty \leq \frac{C_\star}{J} \|\xi - \xi'\|_\infty;$$

(iii) throughout the same neighborhood,

$$\max_{k \in \mathcal{J}_+} \sum_{j \in \mathcal{J}_+} |\partial_{\xi_k} \log \sigma_j(x, \xi; \theta)| \leq C_\star, \qquad \max_{k \in \mathcal{J}_+} |\partial_{\xi_k} \log \sigma_0(x, \xi; \theta)| \leq \frac{C_\star}{J};$$

(iv) the regularized likelihood satisfies

$$\inf_{x \in \mathcal{X}^J,\, s \in \mathcal{S}_J^+, \theta \in \Theta} \left[\sup_{\|\xi - \xi^{\mathrm{BLP}}(x,\theta,s)\|_\infty < r_0} L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}) - \sup_{\|\xi - \xi^{\mathrm{BLP}}(x,\theta,s)\|_\infty \geq r_0} L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}) \right] > 0.$$

Assumption 5(i) bounds the BLP inverse. Assumptions 5(ii) and 5(iii) control local curvature and the score contribution of share errors. Assumption 5(iv) places every global regularized maximizer in the neighborhood where these bounds apply. It is imposed at the

chosen regularization parameter. Under the conditions stated in Appendix B, they hold for simple logit, random-coefficients logit, and random-coefficients nested logit.

Assumption 6. *The regularization parameter satisfies* $\lambda_{N,J} \leq \frac{r_0}{4C_\star^2 J(1+\log J)}$.

When each share σ_j, including the outside share σ_0, is log-concave in ξ, the regularized likelihood is strictly concave and has a unique global maximizer.

Lemma 1 (Log-concavity and global separation)**.** *Suppose Assumptions 1–3 hold and* $\xi \mapsto \log \sigma_j(x, \xi; \theta)$ *is concave for every* $j \in \mathcal{J}$*. Then* L_s^{R} *has a unique global maximizer for every* $s \in \Delta^J$ *and* $\lambda_{N,J} > 0$*. Under Assumptions 5(i)–5(iii) and 6, Assumption 5(iv) also holds.*

Log-concavity holds for simple logit and for the random-coefficients logit and nested-logit models with log-concave taste densities considered in Appendix B, including Gaussian tastes. For lognormal tastes and finite mixtures, Assumption 5(iv) also holds under the conditions of Proposition 3.

Theorem 1 (Uniform finite-sample recovery)**.** *Suppose Assumptions 1–3 and 5–6 hold. Then there exist constants* $c_1, C_1 > 0$*, depending only on* $c, C, c_0, C_\star, r_0$*, such that*

$$\sup_{s^0 \in \mathcal{S}_J} \mathbb{E}\left[\sup_{x \in \mathcal{X}^J,\, \theta \in \Theta,\, \widetilde{\xi} \in \mathcal{A}_N(x,\theta,\hat{s})} \frac{1}{C_1 J \log J} \|\widetilde{\xi} - \xi^{\mathrm{BLP}}(x, \theta, s^0)\|_2^2 \;\middle|\; N, s^0 \right]$$
$$\leq \frac{\frac{J}{N} + \frac{J^2 \log J}{N^2} + J^2 \log J \lambda_{N,J}^2}{(c_1 + J\lambda_{N,J})^2} + \left(\log J + \frac{1}{J\lambda_{N,J}}\right)\left(J \exp\left(-\frac{c_1 N}{J}\right) + \exp(-c_1 N)\right).$$

where the expectation is taken with respect to $Q \sim \mathrm{Multinomial}(N; s^0)$.

The first term on the right-hand side bounds the recovery error when $\hat{s}$ is sufficiently close to s^0. The second term bounds the contribution from the complementary event, which includes sampling zeros. When $\frac{N}{J}$ is large, we can choose $\lambda_{N,J}$ to minimize the first term, which gives $\lambda_{N,J} \asymp \frac{1}{N \log J} + \frac{J}{N^2}$. We call this the *rate-balancing scale*. It is also related to the expected fraction of zero counts. Under multinomial sampling,

$$\mathbb{E}\left[\frac{1}{J} \sum_{j \in \mathcal{J}_+} \mathbf{1}\{Q_j = 0\} \;\middle|\; N, s^0 \right] = \frac{1}{J} \sum_{j \in \mathcal{J}_+} (1 - s_j^0)^N.$$

For $s^0 \in \mathcal{S}_J$ and $\frac{C}{J} \leq \frac{1}{2}$, this expectation lies between $\exp\left(-\frac{2CN}{J}\right)$ and $\exp\left(-\frac{cN}{J}\right)$. Thus both the frequency of sampling zeros and the rate-balancing choice depend on $\frac{N}{J}$. Holding J fixed, a smaller N raises the expected fraction of zeros and the expression $\frac{1}{N \log J} + \frac{J}{N^2}$, giving stronger regularization through a larger $\lambda_{N,J}$.

For asymptotic convergence, we impose the following rates on the sequence (N, J) and the regularization parameter $\lambda_{N,J}$.

Assumption 7 (Asymptotic rates)**.** *Along the sequence* (N, J)*,*

$$\frac{J \log J}{N} \to 0, \qquad J \log J \, \lambda_{N,J} \to 0.$$

The first condition ensures uniform concentration of empirical shares and rules out sampling zeros with probability approaching one. The second makes the regularization error vanish. Under the first condition, we can choose $\lambda_{N,J} \asymp \frac{1}{N}$.

Corollary 1 (Uniform convergence)**.** *Suppose Assumptions 1–3 and 5–7 hold. Then, uniformly over* $s^0 \in \mathcal{S}_J$*,* $L^{\mathrm{R}}_{\hat{s}}(x, \xi; \theta, \lambda_{N,J})$ *has a unique global maximizer in* ξ*, simultaneously for all* $x \in \mathcal{X}^J$ *and* $\theta \in \Theta$*, with probability approaching one. Moreover,*

$$\sup_{x \in \mathcal{X}^J, \, \theta \in \Theta} \|\xi^{\mathrm{R}}_N(x, \theta, \hat{s}) - \xi^{\mathrm{BLP}}(x, \theta, s^0)\|_\infty = O_p\left(\sqrt{\frac{J \log J}{N}} + J \log J \lambda_{N,J}\right).$$

3 Asymptotic Properties

There are T markets. Market t contains J_t inside products, indexed by $\mathcal{J}_{t,+} := \{1, \ldots, J_t\}$, and N_t sampled consumers. The markets form a triangular array, with $J_t = J_{t,T}$ and $N_t = N_{t,T}$. We suppress the additional T index throughout. Write $M_T := \sum_{t=1}^T J_t$, $J_{\max,T} := \max_{t \le T} J_t$, and $N_{\min,T} := \min_{t \le T} N_t$. Let s^0_t and $\hat{s}_t$ denote the true and observed share vectors, respectively. Let z_{jt} denote the instrument vector and $\lambda_{N_t,J_t} > 0$ the regularization parameter used in market t. Fix an open bounded convex neighborhood $\mathcal{N}_0$ of θ_0 whose closure is contained in $\mathrm{int}(\Theta)$.

Assumption 8. *The following conditions hold.*

(i) *For each* t*,* $x_t \in \mathcal{X}^{J_t}$ *and* $Q_t \sim \mathrm{Multinomial}(N_t, s^0_t)$*.*

(ii) *For some* $\delta > 0$ *and* $C_z < \infty$*,* $\sup_T \max_{t \le T, \, j \in \mathcal{J}_{t,+}} \mathbb{E}\|z_{jt}\|_2^{2+\delta} \le C_z$*.*

(iii) $\Pr(s^0_t \in \mathcal{S}_{J_t}$ *for every* $t \le T) \longrightarrow 1$ *as* $T \to \infty$*.*

(iv) $\frac{J_{\max,T} \log(T J_{\max,T})}{N_{\min,T}} \to 0$*.*

Assumption 9. *The following conditions hold.*

(i) *Assumptions 1–5 hold for every market (t,T), with constants $(c_0, c, C, C_\star, r_0)$ that do not depend on t or T.*

(ii) *Let $C_s(x,\xi;\theta) := \partial^2_{\xi\theta'} L_s(x,\xi;\theta)$. Uniformly over $x \in \mathcal{X}^J$, $s \in \mathcal{S}_J^+$, $\theta \in \mathcal{N}_0$, and ξ, ξ' in the r_0-neighborhood of $\xi^{\mathrm{BLP}}(x,\theta,s)$,*

$$\max_{j\in\mathcal{J}_+} \|[C_s(x,\xi;\theta)]_{j\cdot}\|_2 \le \frac{C_\star}{J}, \quad \max_{j\in\mathcal{J}_+} \|[C_s(x,\xi;\theta) - C_s(x,\xi';\theta)]_{j\cdot}\|_2 \le \frac{C_\star}{J}\|\xi - \xi'\|_\infty.$$

Assumption 10. *For some constants $0 < c_\lambda \le C_\lambda < \infty$, one of the following holds uniformly over $t \le T$ and T:*

(i) $\frac{c_\lambda}{N_t} \le \lambda_{N_t,J_t} \le \frac{C_\lambda}{N_t}$;

(ii) $c_\lambda\left(\frac{1}{N_t \log J_t} + \frac{J_t}{N_t^2}\right) \le \lambda_{N_t,J_t} \le C_\lambda\left(\frac{1}{N_t \log J_t} + \frac{J_t}{N_t^2}\right)$.

Let $\mathcal{E}_T := \{s_t^0 \in \mathcal{S}_{J_t},\ \hat{s}_t \in \mathcal{S}_{J_t}^+ \text{ for every } t \le T\}$. By Lemma 5, $\Pr(\mathcal{E}_T) \to 1$. For all sufficiently large T, Lemma 3 implies that $\mathcal{A}_{N_t}(x_t,\theta,\hat{s}_t)$ is a singleton on $\mathcal{E}_T$ for every market and every $\theta \in \Theta$. Lemma 1 shows that under log-concave market share functions, $\mathcal{A}_{N_t}(x_t,\theta,\hat{s}_t)$ is a singleton for every $\theta \in \Theta$ and every $t \le T$. In either case, define

$$\widehat{G}_T^{\mathrm{R}}(\theta) := \frac{1}{M_T}\sum_{t=1}^{T}\sum_{j\in\mathcal{J}_{t,+}} z_{jt}\,\xi^{\mathrm{R}}_{j,N_t}(x_t,\theta,\hat{s}_t)$$

and, for a positive-semidefinite weighting matrix W_T,

$$\widehat{\theta}_T^{\mathrm{R}} := \operatorname*{argmin}_{\theta\in\Theta} \widehat{G}_T^{\mathrm{R}}(\theta)' W_T \widehat{G}_T^{\mathrm{R}}(\theta). \tag{3}$$

On $\mathcal{E}_T$, we define the feasible BLP moment function

$$G_T^{\mathrm{BLP}}(\theta) := \frac{1}{M_T}\sum_{t=1}^{T}\sum_{j\in\mathcal{J}_{t,+}} z_{jt}\xi_j^{\mathrm{BLP}}(x_t,\theta,\hat{s}_t)$$

and let $\widehat{\theta}_T^{\mathrm{BLP}}$ be the corresponding GMM estimator.

Assumption 11. *The following conditions hold.*

(i) $W_T \xrightarrow{p} W$ *for a positive-definite matrix W.*

(ii) *There is a deterministic continuous function $Q : \Theta \to \mathbb{R}$ such that*

$$\sup_{\theta\in\Theta}\left|G_T^{\mathrm{BLP}}(\theta)' W_T G_T^{\mathrm{BLP}}(\theta) - Q(\theta)\right| = o_p(1).$$

(iii) $Q(\theta)$ *has a unique minimizer at* θ_0.

(iv) *There is a deterministic continuous matrix-valued function* Γ *on* $\mathcal{N}_0$ *such that*

$$\sup_{\theta \in \mathcal{N}_0} \|\partial_\theta G_T^{\text{BLP}}(\theta) - \Gamma(\theta)\|_2 = o_p(1), \qquad \sqrt{T}\, G_T^{\text{BLP}}(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Omega).$$

The matrix Ω *is finite and positive semidefinite, and* $\Gamma_0 := \Gamma(\theta_0)$ *has full column rank.*

The moment and Jacobian conditions follow the framework of Theorem 3.2 of Newey and McFadden (1994). Those conditions are used to establish the consistency and asymptotic normality of $\widehat{\theta}_T^{\text{BLP}}$. The following theorems show that $\widehat{\theta}_T^{\text{R}}$ is consistent and first-order asymptotically equivalent to $\widehat{\theta}_T^{\text{BLP}}$.

Theorem 2 (Consistency)**.** *Suppose Assumptions 8–10 and 11(i)–11(iii) hold. Then,* $\widehat{\theta}_T^{\text{R}} \xrightarrow{p} \theta_0$.

Theorem 3 (First-order equivalence)**.** *Suppose Assumptions 8–9 and 11 hold. Assume either*

(i) *Assumption 10(i) holds and* $\sqrt{T}\, \frac{J_{\max,T} \log J_{\max,T}}{N_{\min,T}} \to 0$*; or*

(ii) *Assumption 10(ii) holds and* $\sqrt{T}\, \frac{J_{\max,T}}{N_{\min,T}} \to 0$.

Then

$$\sqrt{T}(\widehat{\theta}_T^{\text{R}} - \widehat{\theta}_T^{\text{BLP}}) = o_p(1), \qquad \sqrt{T}(\widehat{\theta}_T^{\text{R}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, (\Gamma_0' W \Gamma_0)^{-1} \Gamma_0' W \Omega W \Gamma_0 (\Gamma_0' W \Gamma_0)^{-1}).$$

Theorem 3 shows that the regularized GMM estimator is first-order asymptotically equivalent to the feasible BLP GMM estimator. Therefore, the inference procedures developed for the feasible BLP GMM estimator can be applied to the regularized GMM estimator (see, for example, Freyberger, 2015).

4 Monte Carlo Simulations

This section evaluates regularized demand inversion (RDI) using Monte Carlo designs based on Gandhi et al. (2023) (GLS hereafter). We compare RDI with the moment-inequality estimator of GLS in designs with exogenous and endogenous product characteristics. We conduct 500 replications for each design.

4.1 Exogenous Demand Shocks

The first experiment follows Section 8.1 of GLS. Consumer i's utility from product j in market t is

$$u_{ijt} = \alpha_0 + \beta_0 x_{jt} + \lambda_0 x_{jt}\nu_i + \xi_{jt} + \varepsilon_{ijt}, \qquad j = 1, \ldots, J,$$

with outside utility $u_{i0t} = \varepsilon_{i0t}$, $\beta_0 = 1$, $\lambda_0 = 0.5$, $\nu_i \sim \mathcal{N}(0,1)$, Type-I extreme-value idiosyncratic shocks, and $x_{jt} = \frac{j}{10} + e_{jt}$, where $e_{jt} \sim \mathcal{N}(0,1)$ is independent of ξ_{jt}. We consider $\alpha_0 \in \{-9, -10, \ldots, -17\}$ to vary the average zero-cell fraction, and use $T \in \{25, 50, 100\}$ markets, $J = 50$ products, and $N = 10{,}000$ sampled consumers per market.

We consider four distributions for ξ_{jt}: Gaussian shocks, $\xi_{jt} \sim \mathcal{N}(0,1)$; Student-$t_5$ shocks; contaminated-normal shocks, $\xi_{jt} \sim 0.9\mathcal{N}(0,1) + 0.1\mathcal{N}(0,9)$; and bimodal shocks, $\xi_{jt} \sim 0.5\mathcal{N}(-1,1) + 0.5\mathcal{N}(1,1)$. In each replication, the observed count vector is drawn independently across markets from

$$Q \sim \text{Multinomial}(N; s_t^0).$$

Estimators. We compare GLS with the regularized GMM estimator in (3). We set $\lambda_{N,J} = \frac{1}{N\tau^2}$ and vary $\tau \in \{0.5, 1, 1.5, 2, 2.5\}$, labeled RDI-0.5 through RDI-2.5. We also use $\tau^2 = \frac{\log J}{1 + \frac{J}{N}\log J}$, corresponding to the rate-balancing scale, labeled Rate-Balancing. With $J = 50$ and $N = 10{,}000$, we have $\tau \approx 1.96$. We choose the same IVs as in GLS. The regularized estimators use $z_{jt} = (1, x_{jt}, x_{jt}^2 - 1, x_{jt}^3 - 3x_{jt})'$, whereas GLS uses the hypercube instrument functions constructed from x_{jt} with $\bar{r}_T = 50$, as in Gandhi et al. (2023).

Results. Figures 1 and 2 report parameter MSE for $\theta = (\alpha, \beta, \lambda)$. Figures 3 and 4 report MSE for the recovered demand shocks. The GLS procedure does not produce point estimates of demand shocks, so it appears only in the parameter-MSE figures. In all figures, the horizontal axis shows the average zero-cell fraction for each value of α_0.

Across all four shock distributions, the gap in parameter MSE between GLS and RDI widens as zeros become more common. At the highest zero fractions, GLS has parameter MSE between about 1.7 and 6, while every RDI specification remains below 0.25. When zeros are less common, the comparison depends on the scale: RDI-0.5 has larger parameter MSE than GLS in some non-Gaussian designs, while the other RDI scales perform better.

The best scale depends on the shock distribution. Intermediate scales perform best under Gaussian shocks, while RDI-2.5 has the lowest parameter MSE in most Student-t_5, contaminated-normal, and bimodal designs.

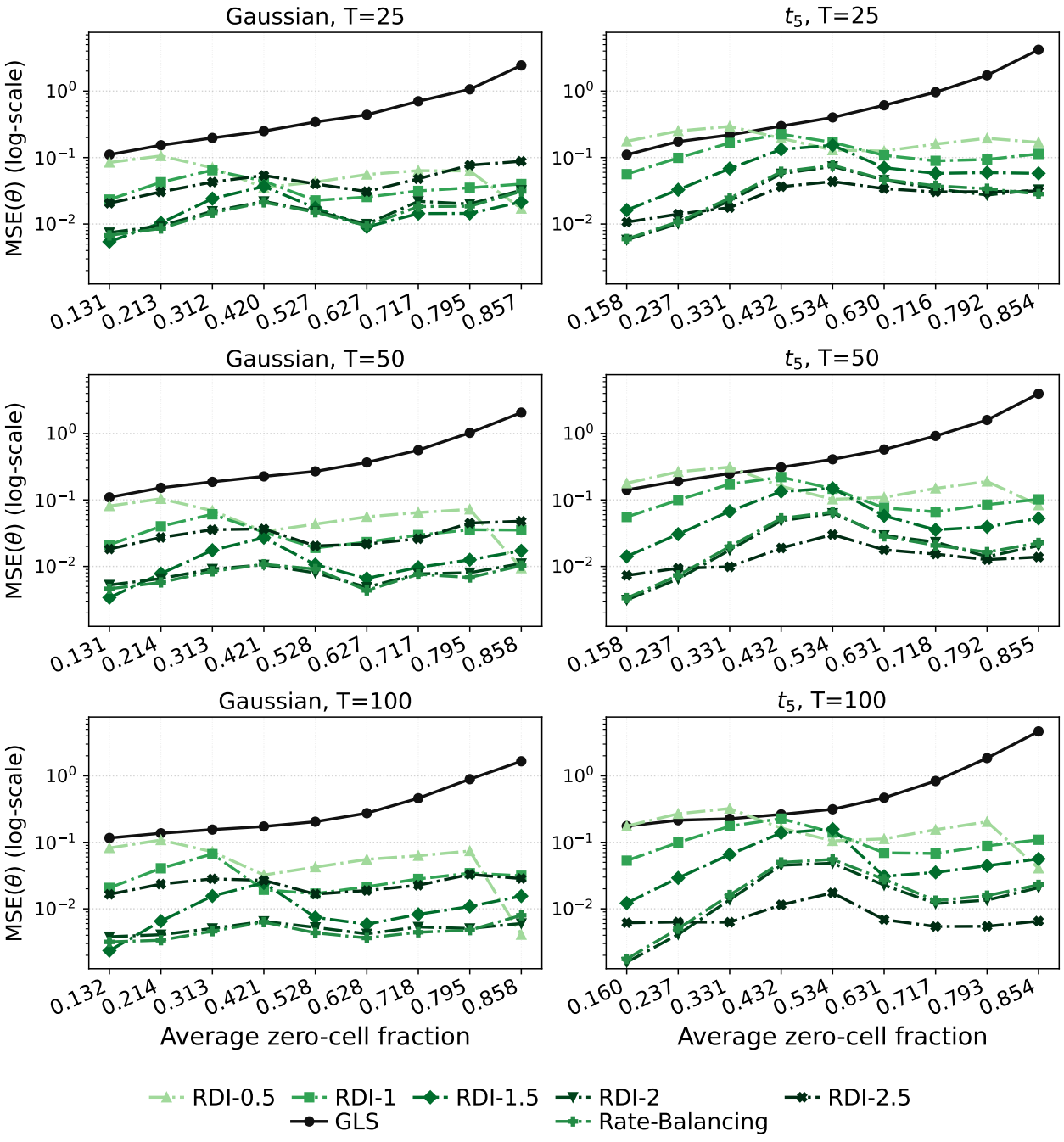


Figure 1: Parameter MSE in the exogenous-ξ design under Gaussian and Student-t_5 shocks.

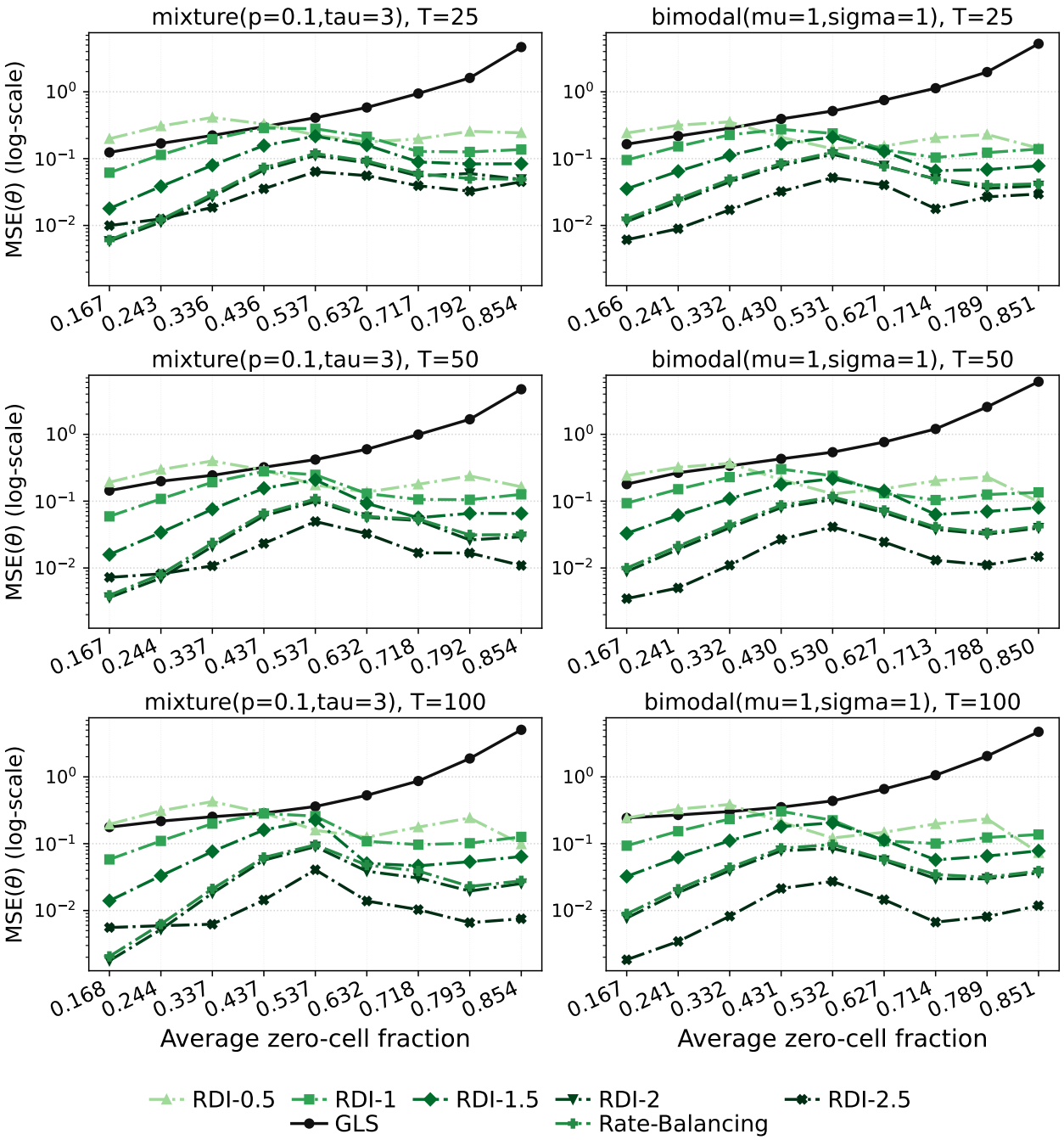


Figure 2: Parameter MSE in the exogenous-ξ design under contaminated-normal and bimodal shocks. Mixture and bimodal refer to the contaminated-normal and bimodal distributions, respectively.

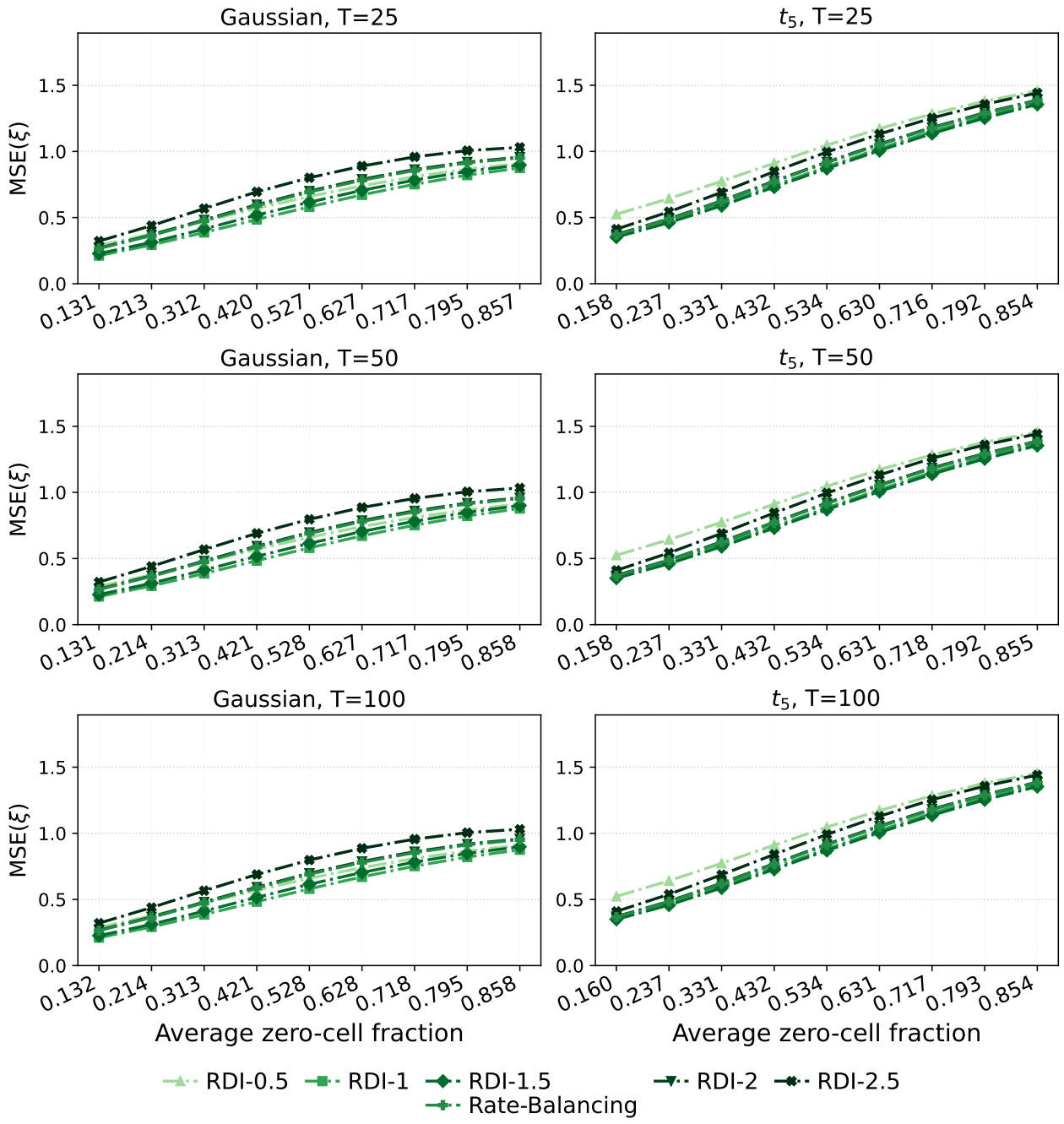


Figure 3: ξ MSE in the exogenous design under Gaussian and Student-t_5 distributions.

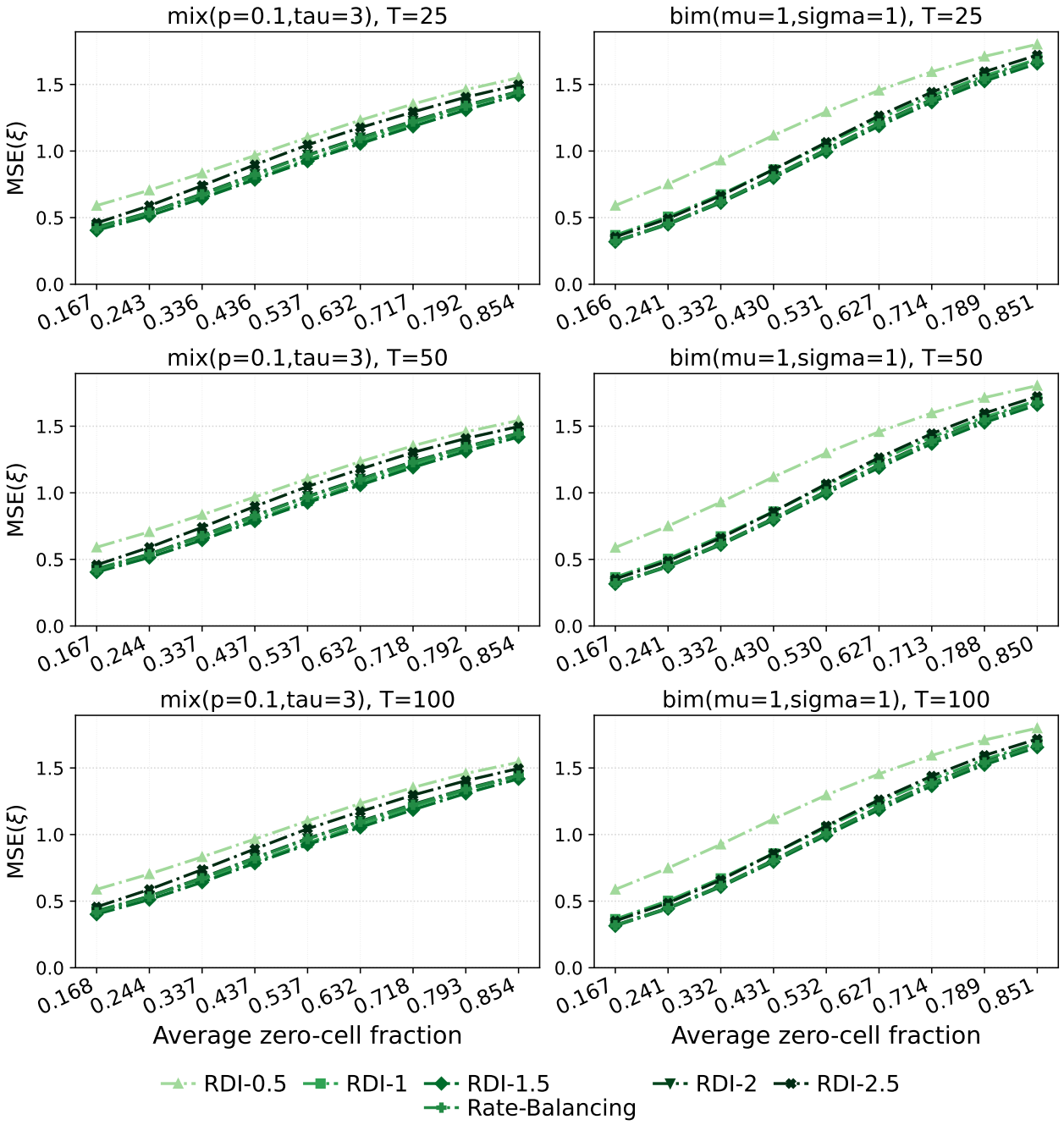


Figure 4: ξ MSE in the exogenous design under contaminated-normal and bimodal distributions. Mixture and bimodal refer to the contaminated-normal and bimodal distributions, respectively.

Increasing the number of markets from 25 to 100 generally improves parameter estimates more for the weaker regularization scales than for the stronger ones. It does not improve the accuracy of the recovered demand shocks. The parameter MSE of GLS changes comparatively little with T.

The ξ MSE rises with the zero-cell fraction and is almost the same for the three values of T. Each demand shock is recovered from a market with the same number of sampled consumers, so adding markets has little effect on its accuracy.

4.2 Endogenous Demand Shocks

The second experiment follows the tuna-market design in Section 8.3 of GLS. Consumer utility is

$$u_{ijt} = \delta_{jt}^{\text{fix}}(\alpha_0) - \nu_i p_{jt} + \xi_{jt} + \varepsilon_{ijt},$$

where $\delta_{jt}^{\text{fix}}(\alpha_0)$ contains the observed tuna-product components, $\log \nu_i \sim \mathcal{N}(\mu_p, \sigma_p^2)$, and $(\mu_p, \sigma_p) = (0, 0.5)$. Following GLS, we hold the coefficients in the fixed utility component at their data-generating values and estimate (μ_p, σ_p). We vary α_0 over six values to change the average zero-cell fraction, and use $T \in \{80, 160\}$ markets. Markets contain about 45 products and, on average, about 17,500 to 17,800 consumers.

The demand shock is heteroskedastic. We draw a base shock $\widetilde{\xi}_{jt}$ from one of the four distributions in Section 4.1 and set

$$\xi_{jt} = \begin{cases} 0.5\widetilde{\xi}_{jt}, & \delta_{jt}^{\text{fix}}(\alpha_0) \geq \text{med}(\delta_{jt}^{\text{fix}}(\alpha_0)), \\ 1.5\widetilde{\xi}_{jt}, & \delta_{jt}^{\text{fix}}(\alpha_0) < \text{med}(\delta_{jt}^{\text{fix}}(\alpha_0)). \end{cases}$$

The price is $p_{jt} = \min\{mc_{jt} + 0.3\exp(\xi_{jt}), p_{\max}\}$, where mc_{jt} is marginal cost and $p_{\max}$ is the maximum observed price in the tuna sample.

Estimators. We compare the same RDI specifications with GLS, setting $\lambda_{N_t,J_t} = \frac{1}{N_t \tau^2}$ in each market. For Rate-Balancing, we compute a common τ by evaluating the preceding formula at the average product count and consumer sample size across markets, which gives $\tau \approx 1.94$. We use the same set of IVs as GLS. Let d_j denote the included UPC-indicator vector and x_{jt}^c the continuous exogenous characteristic. The regularized GMM estimators use $z_{jt} = (1, d_j', mc_{jt}, x_{jt}^c)'$ and one-step GMM. For GLS, we use Bound-$\mathcal{G}_4$ in their notation, which constructs hypercube instrument functions from (mc_{jt}, x_{jt}^c) and their interactions with the UPC indicators.

Results. Figure 5 reports MSE for (μ_p, σ_p). Figures 6 and 7 report MSE for the recovered demand shocks and own-price elasticities, respectively. Elasticities are evaluated at the estimated parameters and the recovered shocks. Because GLS does not recover point estimates of demand shocks, it is not included in the latter two comparisons.

When few cells are zero, RDI and GLS have similar parameter MSE under Gaussian, Student-t_5, and contaminated-normal shocks. GLS performs better under bimodal shocks. At the highest zero fractions, the parameter MSE of GLS is about 1.2 to 1.6, while that of RDI-1 remains below about 0.035.

Stronger regularization performs better in the endogenous design when zeros are common. RDI-1 has the lowest parameter MSE in most designs with more than about 40 percent zero cells, while weaker regularization often performs better when zeros are rare. Rate-Balancing and RDI-2 again have similar parameter MSE.

Increasing from $T = 80$ to $T = 160$ roughly halves parameter MSE for most specifications when zeros are rare. When zeros are common, the improvement remains substantial for RDI-1 but is smaller for the weaker regularization scales and GLS.

The ξ MSE rises with the zero-cell fraction and is nearly the same for $T = 80$ and $T = 160$. The ordering across scales reverses as zeros become more common: RDI-0.5 has the largest ξ MSE when zeros are rare, whereas RDI-2.5 has the largest when zeros are common. Stronger regularization therefore becomes more useful for demand-shock recovery in the sparse designs.

The elasticity results show a similar pattern. Weaker regularization gives lower elasticity MSE when zeros are rare, while RDI-0.5 and RDI-1 perform better as zeros become common.

5 Empirical Application

This section applies our regularized demand inversion estimator to the London beer market. We first document the prevalence of sampling zeros in household purchase data, then estimate demand and compare the implied elasticities, and finally use the estimates to study an alcohol-duty counterfactual.

5.1 Data

The raw data come from Worldpanel by Numerator's Take Home Purchase Panel, a representative panel of households in Great Britain that records grocery purchases brought

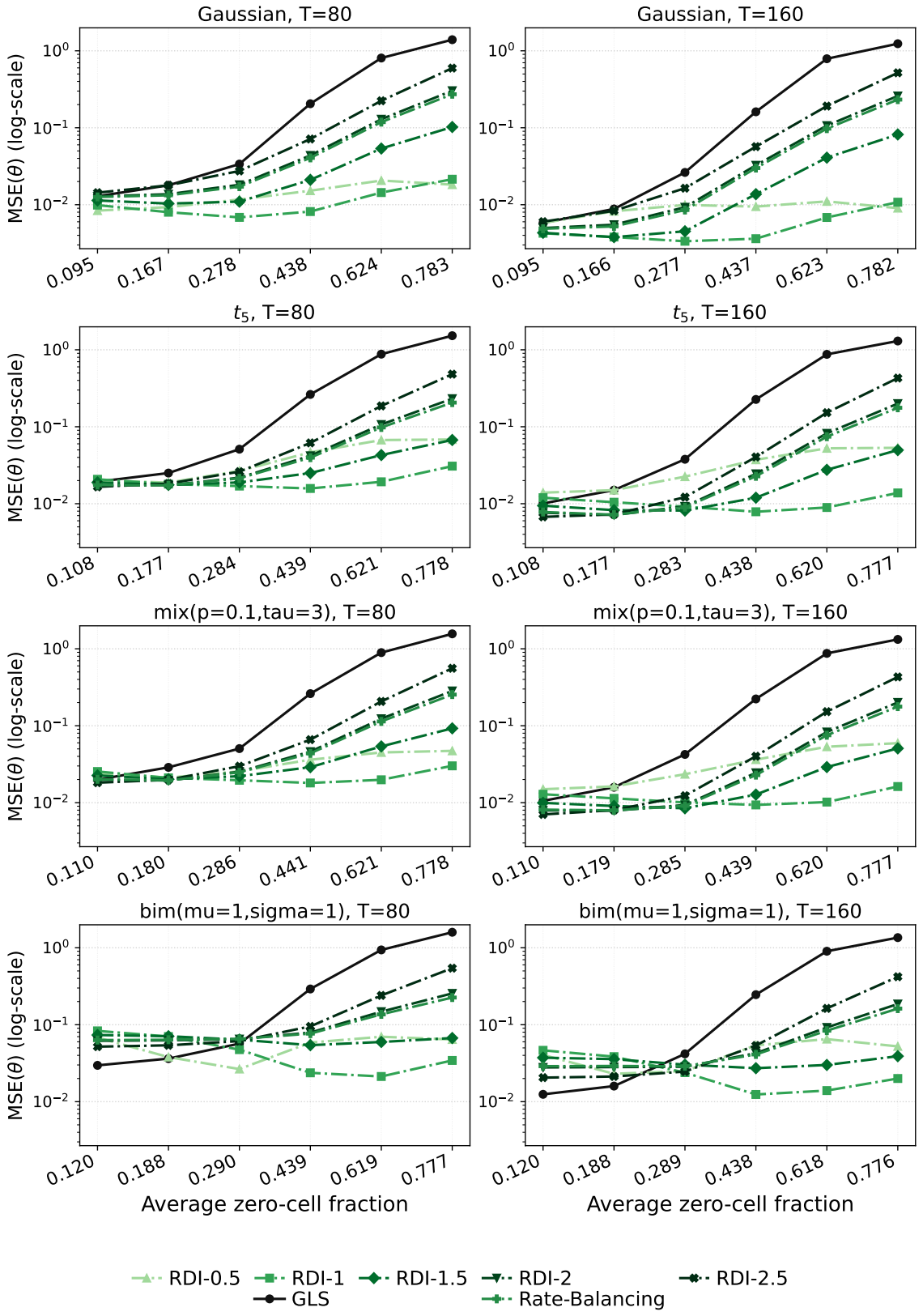


Figure 5: Parameter MSE in the endogenous-ξ design. Mixture and bimodal refer to the contaminated-normal and bimodal distributions, respectively.

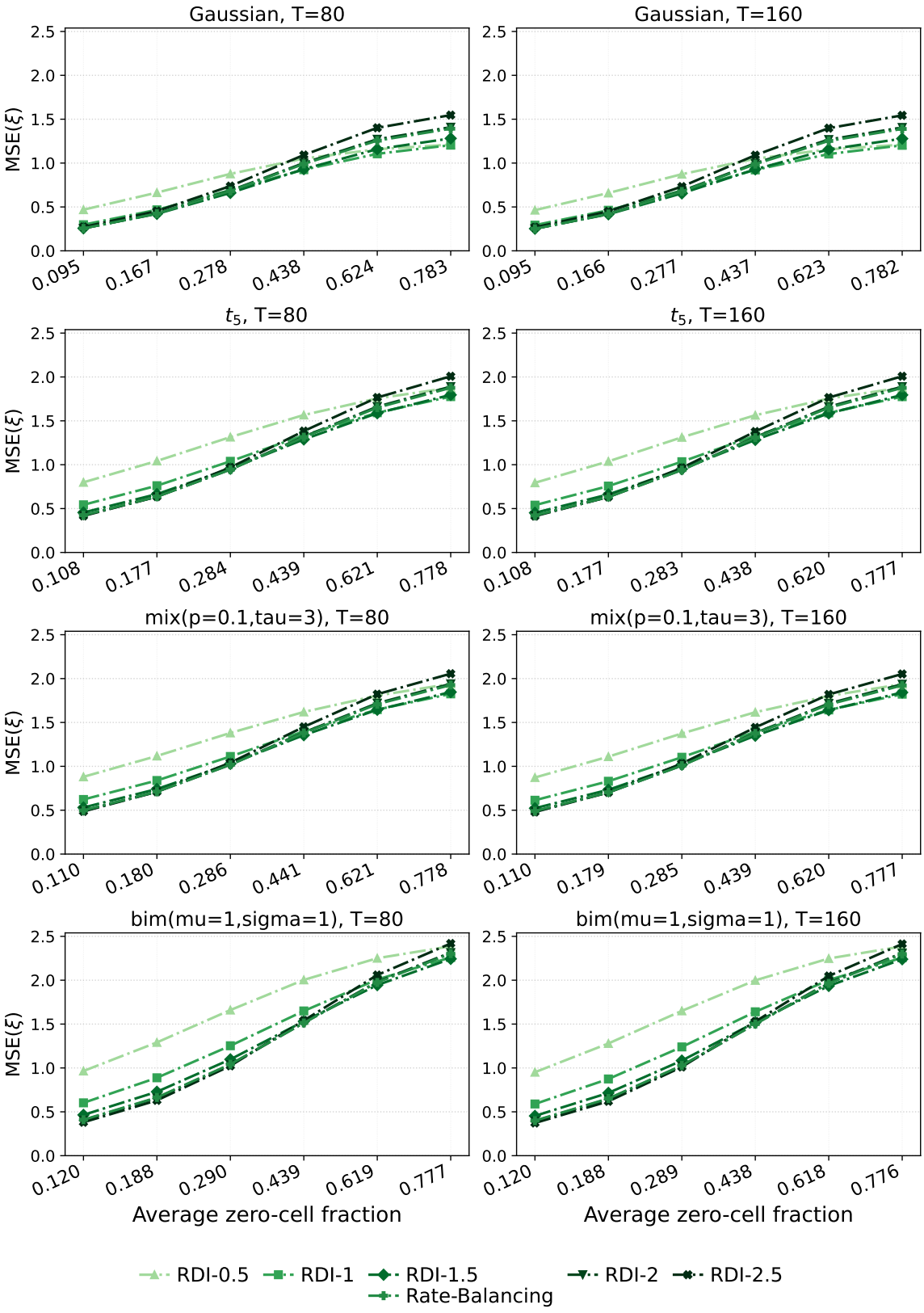


Figure 6: ξ MSE in the endogenous design. Mixture and bimodal refer to the contaminated-normal and bimodal distributions, respectively.

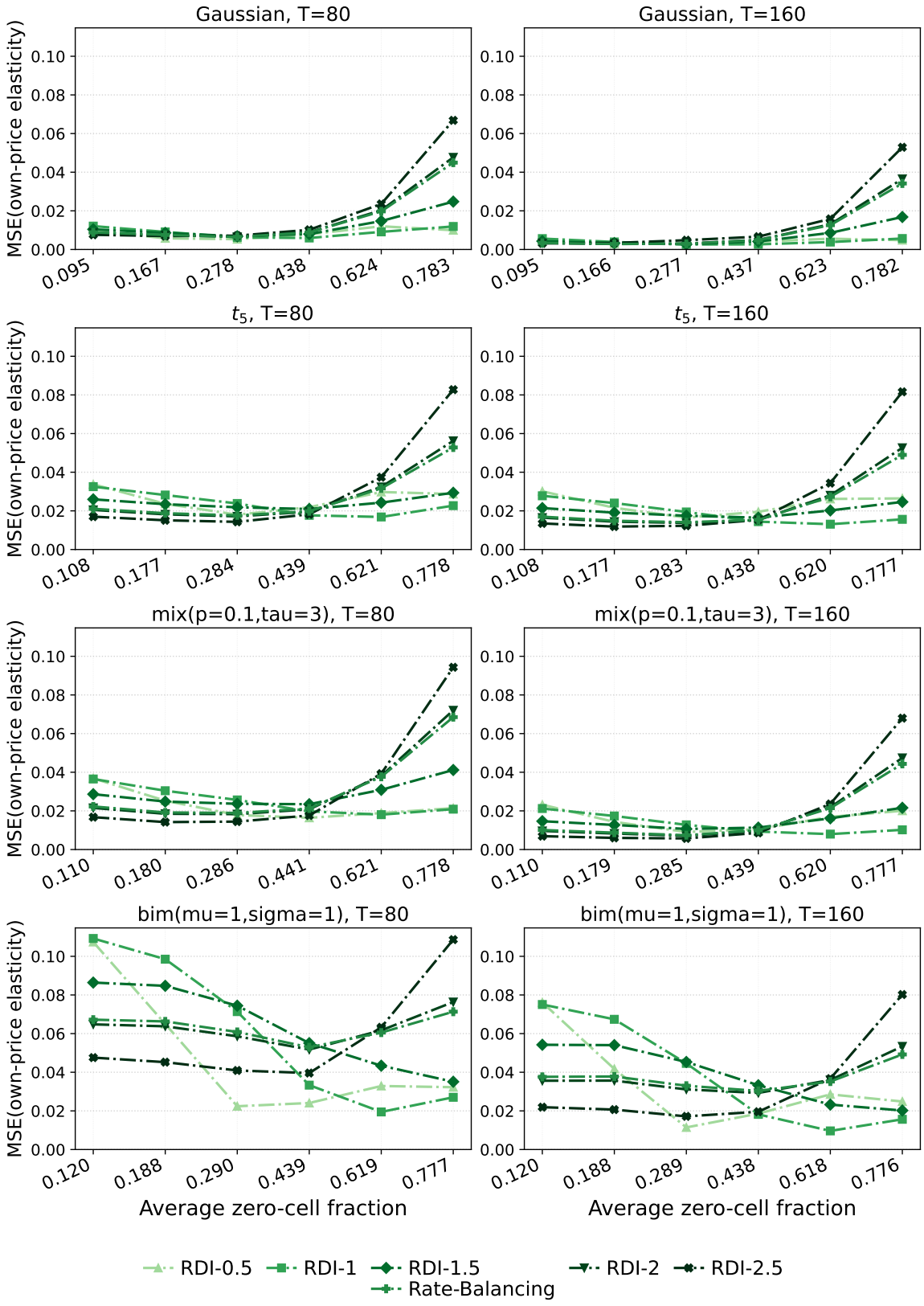


Figure 7: Own-price elasticity MSE in the endogenous-ξ design. Mixture and bimodal refer to the contaminated-normal and bimodal distributions, respectively.

home. We select observations from households resident in London to construct a "London Beer Purchase Panel". The panel contains detailed information on product characteristics, prices, and quantities for each purchase occasion. We use the 2015–2019 panel, focus on the London market, and define a market as a calendar week indexed by t in year y.

First, we highlight the prevalence of zeros in the data. We examine beer, cola, flavored soft drinks, bottled water, instant coffee, tea and infusions, yogurt, and butter and spreadable butter. For each category, a product is a brand–pack-size-bin combination. Table 1 describes the pack-size bins used to define products. We define the market size as the number of households in the panel that bought at least one pack in a year. We further restrict a product to be available in a year if it is purchased at least ten times in that year. Table 2 shows that average product counts range from 83 to 335 across categories, average household counts range from 1,021 to 2,107, and the zero-cell share ranges from 33.3% to 67.0%.

Table 1: Pack-Size Bins for Liquid and Solid Products

Bin label	Pack-size range	Representative formats
Panel A: Liquid products, per-pack volume in liters		
`small_single`	[0.00, 0.40)	275, 330, 355, 375 ml singles
`med_single`	[0.40, 0.60)	440, 500, 550 ml
`large_single`	[0.60, 0.80)	620, 660, 750 ml
`xl_single`	[0.80, 1.10)	990, 1000 ml
`4pack`	[1.10, 2.10)	4×330, 4×440, 4×500
`6_8pack`	[2.10, 3.50)	6×330, 8×330
`10_12pack`	[3.50, 5.00)	10×330, 12×330
`15_18pack`	[5.00, 6.50)	15×330, 18×330
`20_24pack`	[6.50, ∞)	20×330, 24×330, 24×440
Panel B: Solid products, per-pack weight in grams		
`under_100g`	[0, 100)	small sachets, snack-size packs
`100_150g`	[100, 150)	100, 125 g packs
`150_250g`	[150, 250)	150, 200, 227 g packs
`250_500g`	[250, 500)	250, 300, 400, 454 g packs
`500g_1kg`	[500, 1000)	500, 750 g packs
`over_1kg`	[1000, ∞)	1 kg and larger packs

Table 3 summarizes the year-specific product count J_y, week count T_y, household count N_y, purchases, and zero-cell share for the beer category. The sample contains 298–319 products and 990–1,075 households per year, with 12,403–15,808 annual purchases. Between 66.5% and 67.6% of product-week cells have no observed purchases.

Table 2: Zeros in London Grocery Categories

Category	Avg. products	Avg. households	Zero-cell share
Beer	309	1,021	67.0%
Cola	114	1,491	37.5%
Flavored soft drinks	335	1,764	54.0%
Bottled water	204	1,596	45.6%
Yogurt	241	2,107	40.1%
Butter and spreadable butter	83	1,610	33.3%
Instant coffee	184	1,474	50.2%
Tea and infusions	153	1,634	52.8%

Note: A market is a calendar week. Products are brand–pack-size-bin combinations. Statistics are annual averages for 2015–2019, using London households in Worldpanel by Numerator's Take Home Purchase Panel.

Table 3: London Beer Panel

Year	Products J_y	Weeks T_y	Households N_y	Purchases	Zero-cell share
2015	319	53	1,075	15,808	67.4%
2016	308	53	1,041	13,794	67.1%
2017	298	52	997	12,403	67.6%
2018	308	52	1,004	13,816	66.5%
2019	313	52	990	13,580	66.6%

5.2 Model

Demand side. We use a random-coefficients nested logit model similar to Miller and Weinberg (2017). Define two nests as $g = 0$ for the outside option and $g = 1$ for all inside products. Household i in market t obtains indirect utility from product j as

$$u_{ijt}(\theta) = \delta_{jt} + \mu_{ijt} + \zeta_{igt} + (1 - \rho)\varepsilon_{ijt}$$

with $u_{i0t} = \varepsilon_{i0t}$ for the outside option. The error ε_{ijt} is i.i.d. Type-I Extreme Value and the nest-level shock ζ_{igt} is distributed such that $\zeta_{igt} + (1 - \rho)\varepsilon_{ijt}$ is itself Type-I Extreme Value. The nesting parameter $\rho \in [0, 1)$ governs within-nest substitution. The mean utility is

$$\delta_{jt} = \beta_0 + \alpha p_{jt} + \beta_{\text{abv}}\text{ABV}_j + \beta_{\text{ale}}D_j^{\text{ale}} + \beta_{\text{stout}}D_j^{\text{stout}} + \beta_{\text{PL}}D_j^{\text{PL}} + \beta_{\text{vol}}\text{vol}_j + \xi_{jt},$$

where ξ_{jt} is the unobserved product-market shock, D_j^{ale} and D_j^{stout} are indicators for Ale and Stout subsectors, D_j^{PL} is an indicator for private-label products, ABV_j is the alcohol

strength, vol_j is the per-pack volume, and p_{jt} is the price.[3] The observed heterogeneity in the price coefficient is captured by $\mu_{ijt} = \Pi \cdot \tilde{D}_i \cdot p_{jt}$ where Π is the parameter, and $\tilde{D}_i$ is the demeaned income variable for household i.[4]

Let $\theta := (\beta_0, \alpha, \beta_{\text{abv}}, \beta_{\text{ale}}, \beta_{\text{stout}}, \beta_{\text{PL}}, \beta_{\text{vol}}, \Pi, \rho)$ collect the demand parameters. Let $D_{it} := \sum_{k=1}^{J_y} \exp\left(\frac{\delta_{kt}+\mu_{ikt}}{1-\rho}\right)$. The model-implied conditional inside-nest share and inside-nest choice probability are

$$s_{j|g,it} = \frac{\exp\left(\frac{\delta_{jt}+\mu_{ijt}}{1-\rho}\right)}{D_{it}}, \qquad s_{g,it} = \frac{D_{it}^{1-\rho}}{1+D_{it}^{1-\rho}}.$$

Therefore, the model-implied market share of product j in market t in year y is

$$s_{jt} = \frac{1}{N_t} \sum_{i=1}^{N_t} s_{j|g,it}\, s_{g,it},$$

with the outside share $s_{0t} = 1 - \sum_{j=1}^{J_t} s_{jt}$.

Supply side. Firms simultaneously choose prices to maximize profits:

$$\sum_{j \in \mathcal{F}_{f,t}} (p_{jt} - mc_{jt}) s_{jt}(p_t, x_t, \xi_t; \theta) \cdot N_t,$$

where x_t collects the observed product characteristics, ξ_t collects the unobserved product-market shocks, and $\mathcal{F}_{f,t}$ is the set of products owned by firm f in market t. Let O_t be the ownership matrix for market t with $O_{t,jk} = 1$ if products j and k are owned by the same firm and 0 otherwise. The Bertrand–Nash first-order condition implies

$$mc_t = p_t + \left(O_t \odot \frac{\partial s_t}{\partial p_t}\right)^{-1} s_t$$

where mc_t and p_t are the vectors of marginal costs and prices, $\odot$ denotes the Hadamard product, and $\frac{\partial s_t}{\partial p_t}$ is the demand Jacobian matrix with (j,k) element $\frac{\partial s_{jt}}{\partial p_{kt}}$.

[3] If the observed purchase count is positive in market t and product j, p_{jt} equals total expenditure on product j in market t divided by the total number of packs purchased. If the product has zero purchases in market t, we impute p_{jt} by forward and backward carrying the nearest non-missing price for the same product within the same year, using the pooled mean price as a fallback.

[4] In the data, we only observe eight income bins. We take the midpoint of each bin, in units of £10,000, as the income variable, assign £85,000 to the top bin of £70,000 and above, and demean it by its mean among beer-buying households in each year.

We assume

$$\log mc_{jt} = \gamma_0 + \gamma_{\text{abv}}\text{ABV}_j + \gamma_{\text{ale}}D_j^{\text{ale}} + \gamma_{\text{stout}}D_j^{\text{stout}} + \gamma_{\text{vol}}\text{vol}_j + \phi_{m(j)} + \omega_{jt},$$

where $\phi_{m(j)}$ is the manufacturer fixed effect, and ω_{jt} is the unobserved cost shock.

5.3 Estimation Results

Prices are potentially correlated with ξ_{jt}. We use four excluded instruments for price. The first is the weekly mean non-London per-litre price of the same product and the second is the weekly mean non-London per-litre price of products in the same range-trading company.[5] The other two are rival-manufacturer sums of ABV and Stout indicators.[6]

Estimators. We compare four estimators. The first is the regularized demand inversion estimator, labeled "RDI", with $\lambda_{N_t,J_t} = \frac{1}{N_t\tau^2}$. The second is the standard BLP estimator computed by dropping all zero cells, labeled "ES." The third is the Laplace-smoothed estimator that uses $\tilde{s}_{jt} = \frac{(Q_{jt}+1)}{(N_t+J_y+1)}$, labeled "LS." These three estimators use two-step GMM. The fourth is the GLS moment-inequality estimator, labeled "GLS." Similar to the Monte Carlo design, we construct hypercube instrument functions for ABV, package volume, and the excluded instruments, using refinement parameter $r = 30$. Reported parameter standard errors use 100 market-level bootstrap replications.

Results. Table 4 presents the demand estimates and the own-price elasticities.[7] The RDI estimates have the expected signs. Holding price fixed, households prefer stronger beer, larger packs, and branded products, and they prefer lagers to ales and stouts. At the mean income, the price coefficient is $\hat{\alpha} = -0.972$ per pound, and each additional £10,000 of income reduces its magnitude by $\widehat{\Pi} = 0.189$. The nesting parameter $\hat{\rho} = 0.584$ implies that households substitute more readily among beers than between beer and the outside option. ES and LS give price coefficients and price–income interactions of similar magnitude but a

[5] Slade (2004); Pinkse and Slade (2004) use similar cross-region price instruments in studies of the UK beer market, exploiting common cost variation across regions.

[6] Armstrong (2016) shows that BLP instruments can be weak when the number of products is large. Although our IVs contain proxy variables for common cost shocks, we also report the first-stage F-statistic, which is $26.44 > 10$.

[7] For GLS, we recover ξ_{jt} by inverting the demand system at the estimated parameters using Laplace-smoothed shares, and use the recovered demand shocks to compute elasticities. ES elasticities are available only for the 26,703 product-market cells with positive purchases, whereas the other estimators use all 81,019 cells.

weaker nesting parameter of about 0.31, whereas GLS gives a much smaller price coefficient of -0.154.

RDI implies a mean own-price elasticity of -3.315 and a median of -3.220, compared with means of -1.202 for ES, -1.291 for LS, and -0.443 for GLS. Only 4 of the 81,019 RDI elasticities exceed -1, compared with 33.0% of the ES cells, 24.6% of the LS cells, and 81.5% of the GLS cells. LS and GLS also imply upward-sloping demand, with positive own-price elasticities in 13.4% and 15.2% of cells. Such elasticities are difficult to reconcile with the Bertrand–Nash pricing model used below. The choice of estimator therefore has quantitatively important consequences for implied price sensitivity.

The gap reflects two features of the estimates. First, holding the price coefficients fixed, the own-price elasticity of a product with a small within-nest share is approximately proportional to $1/(1-\rho)$, which equals 2.41 under RDI and 1.46 under ES and LS. Second, ES, LS, and GLS imply that households in the top income bin are almost insensitive to price. Averaged over years, the implied price coefficient $\hat{\alpha} + \widehat{\Pi}\tilde{D}_i$ for the top bin is -0.19 under RDI, -0.02 under ES, 0.04 under LS, and 0.08 under GLS, whereas for the lowest bin it is -1.70, -1.66, -1.58, and -0.37. This difference matters most for expensive products. Under the RDI estimates, households in the top bin make 93% of the predicted purchases of products in the top price tercile, above £3.29 per pack. In that tercile, 84.0% of the ES elasticities and 72.3% of the LS elasticities exceed -1, compared with none of the RDI elasticities. Restricting RDI to the positive-purchase cells used by ES gives a mean elasticity of -3.337, close to its full-sample mean of -3.315.

For comparison, Slade (2004) reports mean own-price elasticities of about -3.6 under nested logit and -4.6 under a distance-metric demand system, while Miller and Weinberg (2017) reports median own-price elasticities between -4.33 and -6.10 across random-coefficients nested-logit specifications. The RDI elasticities are of the same order as these estimates, whereas the ES, LS, and GLS elasticities are much smaller in magnitude. We use RDI, shown in the first column, for the supply-side inversion and counterfactual analysis.

Figures 8 and 9 show the income patterns implied by the RDI estimates. The positive price–income interaction, $\widehat{\Pi} = 0.189$, implies that higher-income households are less price sensitive. Mean income-specific own-price elasticities range from about -12.32 in the lowest income bin to -1.58 in the highest. Predicted beer purchase probabilities rise from 8.1% to 62.1% across the same income bins. Consequently, households in the top bin, with annual income of at least £70,000, make up 18.1% of beer-buying households but account for 43.9% of predicted purchases.

Table 4: Demand Estimates and Own-Price Elasticities

	RDI	ES	LS	GLS
Panel A: Parameter estimates				
β_0	-3.056 (0.915)	-3.065 (0.002)	-3.036 (0.157)	-1.466 (0.114)
α	-0.972 (0.181)	-0.863 (0.010)	-0.800 (0.333)	-0.154 (0.020)
β_{ABV}	0.114 (0.030)	0.048 (0.006)	0.059 (0.017)	0.035 (0.017)
β_{ale}	-0.359 (0.059)	-0.307 (0.009)	-0.329 (0.132)	-0.149 (0.058)
β_{stout}	-0.081 (0.024)	-0.049 (0.019)	-0.039 (0.011)	-0.167 (0.084)
β_{PL}	-0.325 (0.050)	-0.195 (0.019)	-0.169 (0.050)	-0.062 (0.076)
β_{vol}	0.495 (0.171)	0.197 (0.019)	0.233 (0.049)	-0.013 (0.030)
Π	0.189 (0.061)	0.205 (0.005)	0.202 (0.092)	0.056 (0.005)
ρ	0.584 (0.105)	0.312 (0.008)	0.316 (0.128)	0.667 (0.012)
Panel B: Own-price elasticities				
Mean	-3.315	-1.202	-1.291	-0.443
Median	-3.220	-1.414	-1.630	-0.856
Standard deviation	0.902	0.546	0.817	1.072
Share above -1	0.0%	33.0%	24.6%	81.5%
Share above 0	0.0%	2.4%	13.4%	15.2%

Note: London beer purchase panel, 2015–2019. RDI, ES, and LS use two-step GMM. GLS uses the Gandhi et al. (2023) estimator. Parentheses report bootstrap standard errors based on 100 market-level replications. RDI uses fixed $\tau = 1.45046$. Elasticities use all 81,019 product-market cells, except for ES, which uses the 26,703 cells with positive purchases.

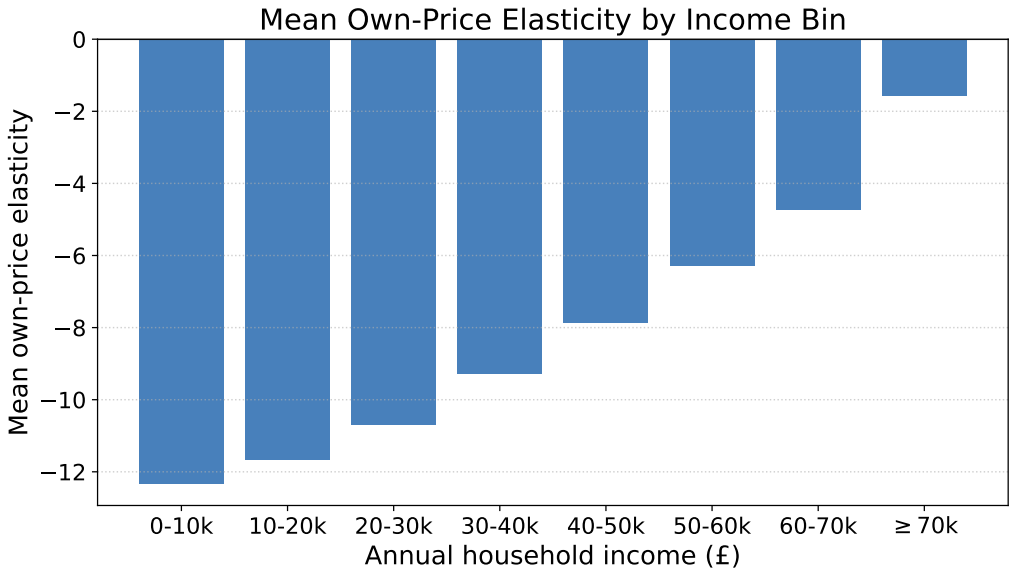


Figure 8: Mean own-price elasticity by income bin.

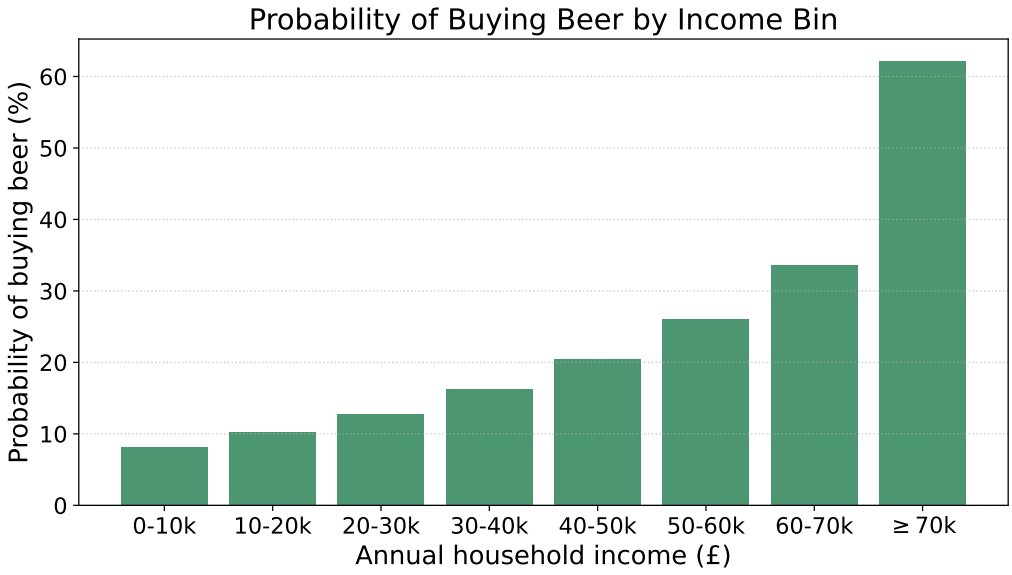


Figure 9: Purchase probability by income bin.

Regularization sensitivity. We use a common $\tau = 1.45046$ as the benchmark, obtained by evaluating $\tau^2 = \frac{\log J}{1+\frac{J}{N}\log J}$ at $J = 300$ and $N = 1{,}000$, following the rate-balancing scale. Larger values of τ imply weaker regularization. Figure 10 plots the mean own-price elasticity against τ over the range 0.25–2.5. The mean moves from -3.60 at $\tau = 0.25$ to -2.96 at 0.75, then becomes more negative, reaching -3.84 at 2.5. At the rate-balancing scale, the mean is -3.315.

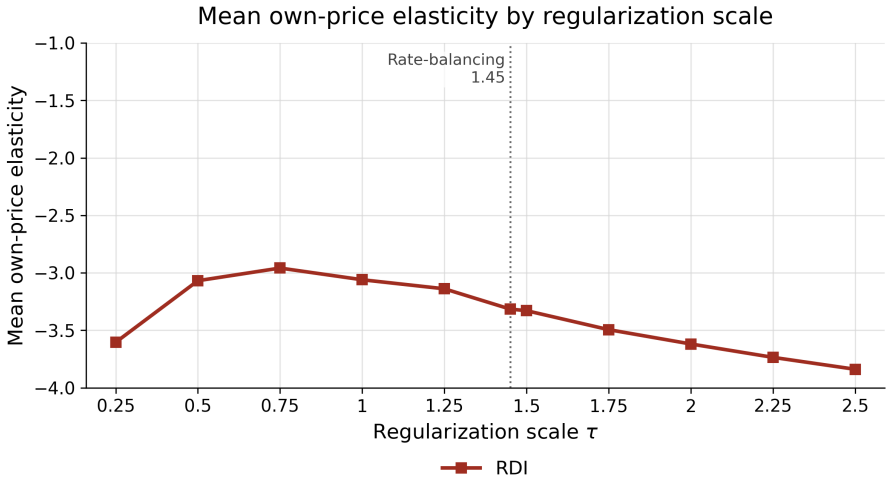


Figure 10: Regularization sensitivity of mean own-price elasticity.

Supply-side results. Table 5 reports the marginal costs and markup rates recovered from the Nash–Bertrand supply-side inversion.

Table 5: Marginal Cost and Markup

	Panel A: Marginal cost mc_{jt} (£/pack)					*Panel B: Markup rate $\frac{p-mc}{p}$*				
Year	Mean	SD	P5	Median	P95	Mean	SD	P5	Median	P95
2015	2.206	2.096	0.614	1.405	7.230	0.332	0.063	0.261	0.311	0.451
2016	2.231	2.432	0.613	1.271	7.610	0.344	0.071	0.260	0.323	0.485
2017	2.481	2.722	0.626	1.342	9.264	0.330	0.068	0.216	0.315	0.451
2018	2.575	2.988	0.601	1.278	9.763	0.330	0.073	0.183	0.323	0.439
2019	2.484	2.844	0.585	1.263	9.714	0.333	0.072	0.183	0.327	0.436
pooled	2.392	2.634	0.606	1.302	8.746	0.334	0.070	0.215	0.321	0.454

Note: London beer purchase panel, 2015–2019. Panel A shows the marginal-cost distribution, Panel B shows the markup-rate distribution. Both panels are summarized per year and pooled across the sample.

The pooled mean marginal cost is £2.39 per pack, with a median of £1.30. Annual mean marginal costs range from £2.21 to £2.57, and the pooled median markup rate is 32.1% of price. The log-marginal cost regression gives an R^2 of 0.84.

5.4 Counterfactual: 2026 Alcohol Duty Schedule

Table 6 reports the alcohol duty rates. We apply the 2026 duty schedule to each sample year by adding the change in duty per pack to marginal cost and solving for new Bertrand–Nash equilibrium prices. Ownership, product characteristics, preferences, and demand shocks are held fixed. Most of the change falls on standard-strength beer. Products with ABV between 3.5% and 7.5%, which account for 93–96% of predicted purchases in each year, face a 2026 rate that is 22.9% above the 2015–2016 rate and 18.3% above the 2017–2019 rate. Products between 2.8% and 3.5% ABV, at most 1.5% of predicted purchases, receive duty cuts of about 46–48%, and the few products at or above 8.5% ABV face increases of 24–28%.

Table 7 reports the counterfactual results. Applying the 2026 duty schedule reduces consumer surplus and profit in every sample year, while pooled duty revenue changes little. On average, consumer surplus falls by £0.090 per household-market, and profit falls by £0.014. Over a year, the consumer-surplus loss is about £4.70 per household. Duty revenue falls by £0.004, a small change relative to its bootstrap standard error of £0.006. The combined change in consumer surplus, profit, and duty revenue is $-$£0.108, or 6.1% of their baseline sum. Consumers bear 83% of this loss, firms 13%, and duty revenue 3%.

Table 6: Beer Duty Schedules

Pounds per litre of product

Panel A: Sample-period schedules

Sample years	$>$ 1.2–2.8% ABV	$>$ 2.8–7.5% ABV	$>$ 7.5% ABV
2015–2016	$0.0810A$	$0.1837A$	$0.2385A$
2017–2019	$0.0842A$	$0.1908A$	$0.2477A$

Panel B: 2026 non-draught schedule

ABV band	Duty
0–1.2%	0
1.3–3.4%	$0.0996A$
3.5–8.4%	$0.2258A$
8.5–22%	$0.3062A$
$>$ 22%	$0.3399A$

Note: A denotes ABV in percentage points; for example, $A = 5$ for a 5% beer. The counterfactual uses the non-draught schedule because the sample covers off-trade purchases. Source: HMRC historical and 2026 alcohol duty schedules.

Table 7: Alcohol Duty Counterfactual

Year	CS	Δ CS	Profit	Δ Profit	Tax revenue	Δ Tax	Δ Total surplus
2015	1.250	-0.114	0.300	-0.014	0.410	-0.000	-0.128
	(0.149)	(0.007)	(0.018)	(0.003)	(0.003)	(0.007)	(0.004)
2016	1.208	-0.109	0.301	-0.016	0.401	0.003	-0.121
	(0.137)	(0.006)	(0.019)	(0.003)	(0.003)	(0.007)	(0.004)
2017	1.073	-0.082	0.273	-0.013	0.386	-0.001	-0.096
	(0.130)	(0.004)	(0.016)	(0.002)	(0.003)	(0.005)	(0.003)
2018	1.040	-0.076	0.264	-0.015	0.387	-0.011	-0.101
	(0.138)	(0.003)	(0.014)	(0.001)	(0.004)	(0.005)	(0.003)
2019	1.003	-0.069	0.258	-0.015	0.364	-0.010	-0.093
	(0.139)	(0.002)	(0.013)	(0.001)	(0.003)	(0.004)	(0.003)
pooled	1.119	-0.090	0.280	-0.014	0.390	-0.004	-0.108
	(0.137)	(0.004)	(0.016)	(0.002)	(0.003)	(0.006)	(0.003)

Note: London beer purchase panel, 2015–2019. The counterfactual applies the 2026 beer duty schedule to each sample year. All entries are pounds per household-market. CS, profit, and tax revenue report baseline levels. Δ denotes the counterfactual minus the baseline. Tax revenue is beer-duty revenue and excludes VAT. Total surplus equals consumer surplus plus profit plus tax revenue. Bootstrap standard errors are in parentheses.

6 Conclusion

We propose regularized demand inversion by placing a quadratic penalty on demand shocks in the multinomial log-likelihood. The regularized inverse exists for every count vector, including those containing zeros. We establish finite-sample recovery bounds and a rate-balancing choice of the regularization parameter. We show that the resulting GMM estimator is consistent and first-order asymptotically equivalent to the feasible BLP estimator. In the London beer application, RDI yields substantially more price-elastic demand than the comparison estimators.

References

Ackerberg, D. A. and Rysman, M. (2005). Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *The RAND Journal of Economics*, 36(4):771–788.

Adam, H., He, P., and Zheng, F. (2024). Machine learning for demand estimation in long tail markets. *Management Science*, 70(8):5040–5065.

Albert, A. and Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1):1–10.

Allen, R. (2026). Exogeneity conditions in markets with finite consumers.

Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, 77(2):489–536.

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

Berry, S., Gandhi, A., and Haile, P. A. (2013). Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S., Linton, O. B., and Pakes, A. (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies*, 71(3):613–654.

Berry, S. and Pakes, A. (2007). The pure characteristics demand model. *International Economic Review*, 48(4):1193–1225.

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

Beyhum, J., Dhaene, G., Pakel, C., and Weidner, M. (2026). Approximate operator inversion for average effects in nonlinear panel models. *arXiv preprint arXiv:2605.05037*.

de Lima, L. G. (2024). Flexible demand estimation and zero market shares. Technical report, Texto para discussão.

Dubé, J.-P., Hortaçsu, A., and Joo, J. (2021). Random-coefficients logit demand estimation with zero-valued market shares. *Marketing Science*, 40(4):637–660.

Dubois, P., Griffith, R., and O'Connell, M. (2020). How well targeted are soda taxes? *American Economic Review*, 110(11):3661–3704.

D'Haultfœuille, X., Durrmeyer, I., and Février, P. (2019). Automobile prices in market equilibrium with unobserved price discrimination. *The Review of Economic Studies*, 86(5):1973–1998.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1):27–38.

Freyberger, J. (2015). Asymptotic theory for differentiated products demand models with many markets. *Journal of Econometrics*, 185(1):162–181.

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

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21(16):2409–2419.

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

Hu, J. (2026). *Three Essays on Demand Estimation With Zero Market Shares*. PhD thesis, The Pennsylvania State University.

King, G. and Zeng, L. (2001). Logistic regression in rare events data. *Political analysis*, 9(2):137–163.

Kosmidis, I. and Firth, D. (2021). Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometrika*, 108(1):71–82.

Li, J. (2023). Compatibility and investment in the us electric vehicle market. *Unpublished manuscript, MIT*.

Lieber, J. and Wiemann, T. (2022). Demand estimation with finitely many consumers. Technical report, Working Paper.

Lu, Z. and Shimizu, K. (2025). Estimating discrete choice demand models with sparse market-product shocks. Staff Working Paper 2025-10, Bank of Canada. arXiv:2501.02381.

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

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing.

Handbook of econometrics, 4:2111–2245.

O'Connell, M. and Smith, K. (2024). Optimal sin taxation and market power. *American Economic Journal: Applied Economics*, 16(4):34–70.

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

Prékopa, A. (1973). On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, 34:335–343.

Quan, T. W. and Williams, K. R. (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics*, 49(4):877–913.

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

Stourm, L. and Stourm, V. (2025). Estimating sparse spatial demand to manage crowdsourced supply in the sharing economy. *Marketing Science*, 44(4):777–801.

Yan, Z., Zhang, Z., Chen, M., Li, J., and Fernández-Val, I. (2026). Robust priors in nonlinear panel models with individual and time effects. *arXiv preprint arXiv:2604.03663*.

A Proofs

A.1 Proofs in Section 2

Proof of Proposition 1. Fix (x, θ). For every $s \in \Delta^J$, the function $-L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J})$ is continuous and coercive because $-L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}) \geq \frac{\lambda_{N,J}}{2}\|\xi\|_2^2$. Hence $\mathcal{A}_N(x, \theta, s)$ is nonempty and compact.

Now let $s \in \mathrm{int}(\Delta^J)$, write $\xi^* = \xi^{\mathrm{BLP}}(x, \theta, s)$, and suppress (x, θ). Then ξ^* is the unique global maximizer of L_s. The nonsingular inside-share Jacobian and global injectivity give a unique continuous inverse near s. For any sequence (ξ_n) such that $L_s(\xi_n) \to L_s(\xi^*)$, Pinsker's inequality gives $\sigma(\xi_n) \to s$, and continuity of the inverse then gives $\xi_n \to \xi^*$. Thus, for every $r > 0$,

$$L_s(\xi^*) - \sup_{\|\xi - \xi^*\|_\infty \geq r} L_s(\xi) > 0. \tag{4}$$

At ξ^*, differentiation using $\sigma(\xi^*) = s$ and $\sum_{j \in \mathcal{J}} \sigma_j(\xi) = 1$ gives

$$-\partial^2_{\xi\xi'} L_s(\xi^*) = \partial_\xi \sigma(\xi^*)' \,\mathrm{diag}(s_0^{-1}, \ldots, s_J^{-1}) \partial_\xi \sigma(\xi^*). \tag{5}$$

The matrix is positive definite because the inside-share Jacobian is nonsingular. By con-

tinuity, the likelihood Hessian remains negative definite on a closed sup-norm ball of some radius $r > 0$ around ξ^*. Therefore, L_s^{R} is strictly concave on this ball.

For any $\widetilde{\xi} \in \mathcal{A}_N(x, \theta, s)$, we have

$$0 \leq L_s(\xi^*) - L_s(\widetilde{\xi}) \leq \frac{\lambda_{N,J}}{2} \|\xi^*\|_2^2.$$

For sufficiently small $\lambda_{N,J} > 0$, (4) implies $\|\widetilde{\xi} - \xi^*\|_\infty < r$, so strict concavity gives uniqueness.

For nearby θ, global maximizers remain in this ball by continuity and the quadratic penalty. The implicit-function theorem then gives local continuous differentiability in θ. □

Lemma 2. *Suppose Assumptions 1–3, 5(ii), and 5(iii) hold. Uniformly over $x \in \mathcal{X}^J$, $s, s' \in \mathcal{S}_J^+$, and $\theta \in \Theta$,*

$$\|\xi^{\mathrm{BLP}}(x, \theta, s') - \xi^{\mathrm{BLP}}(x, \theta, s)\|_\infty \leq C_\star^2 \left(J \max_{j \in \mathcal{J}_+} |s'_j - s_j| + |s'_0 - s_0| \right).$$

Proof. The set $\mathcal{S}_J^+$ is convex, so $s(u) = s + u(s' - s)$ belongs to it for $0 \leq u \leq 1$. The inside-share Jacobian is nonsingular by Assumption 2, and the inverse-function theorem makes $u \mapsto \xi^{\mathrm{BLP}}(x, \theta, s(u))$ continuously differentiable. Differentiating the likelihood first-order condition (FOC) at $\xi^{\mathrm{BLP}}(x, \theta, s(u))$ with respect to u gives

$$H_{s(u)}\big(x, \xi^{\mathrm{BLP}}(x, \theta, s(u)); \theta\big) \frac{d}{du} \xi^{\mathrm{BLP}}(x, \theta, s(u)) = \sum_{j \in \mathcal{J}} (s'_j - s_j) \partial_\xi \log \sigma_j\big(x, \xi^{\mathrm{BLP}}(x, \theta, s(u)); \theta\big).$$

The maximum norm of the RHS is at most $C_\star \left(\max_{j \in \mathcal{J}_+} |s'_j - s_j| + \frac{|s'_0 - s_0|}{J} \right)$. The inverse information matrix has norm at most $C_\star J$. Multiply the two bounds and note that $\xi^{\mathrm{BLP}}(x, \theta, s') - \xi^{\mathrm{BLP}}(x, \theta, s) = \int_0^1 \frac{d}{du} \xi^{\mathrm{BLP}}(x, \theta, s(u))\, du$. □

Lemma 3 (Local regularized inverse)**.** *Suppose Assumptions 1–3, 5(i)–5(iii), and 6 hold. For $x \in \mathcal{X}^J$, $s \in \mathcal{S}_J^+$, and $\theta \in \Theta$, let $\mathcal{B} = \{\xi \in \mathbb{R}^J : \|\xi - \xi^{\mathrm{BLP}}(x, \theta, s)\|_\infty \leq r_0\}$. The equation $\partial_\xi L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}) = 0$ has a unique solution $\xi_L \in \mathcal{B}$, which uniquely maximizes L_s^{R} on $\mathcal{B}$ and satisfies*

$$\|\xi_L - \xi^{\mathrm{BLP}}(x, \theta, s)\|_\infty \leq 2C_\star^2 J \log J \lambda_{N,J} \leq \frac{r_0}{2}. \tag{6}$$

For $\xi \in \mathcal{B}$,

$$H_s(x, \xi; \theta) \succeq \frac{1}{2C_\star J} I_J, \qquad \big\|(H_s(x, \xi; \theta) + \lambda_{N,J} I_J)^{-1}\big\|_\infty \leq 2C_\star J. \tag{7}$$

Under Assumption 5(iv), ξ_L is also the unique maximizer of L_s^{R} on $\mathbb{R}^J$.

Proof. Fix (x, s, θ) and write $\xi^* = \xi^{\mathrm{BLP}}(x, \theta, s)$. Then $\partial_\xi L_s(x, \xi^*; \theta) = 0$ and $\partial^2_{\xi\xi'} L_s(x, \xi; \theta) = -H_s(x, \xi; \theta)$. By (5), $H_s(\xi^*)$ is symmetric and positive definite, so

$$\frac{1}{\lambda_{\min}(H_s(\xi^*))} = \|H_s(\xi^*)^{-1}\|_2 \le \|H_s(\xi^*)^{-1}\|_\infty \le C_\star J.$$

Thus $\lambda_{\min}(H_s(\xi^*)) \ge \frac{1}{C_\star J}$. The Hessian-Lipschitz bound and $r_0 < \frac{1}{4C_\star^2}$ give $H_s(\xi) \succeq \frac{1}{2C_\star J} I_J$ on $\mathcal{B}$.

Define the map

$$\Phi(\xi) = \xi + H_s(\xi^*)^{-1} \left(\partial_\xi L_s(x, \xi; \theta) - \lambda_{N,J} \xi\right).$$

Since $H_s(\xi^*)$ is invertible, $\Phi(\xi) = \xi$ if and only if $\partial_\xi L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}) = 0$.

The Jacobian is $\partial_\xi \Phi(\xi) = -H_s(\xi^*)^{-1}(H_s(\xi) - H_s(\xi^*) + \lambda_{N,J} I_J)$. By Assumption 6, for $\xi \in \mathcal{B}$,

$$\|\partial_\xi \Phi(\xi)\|_\infty \le C_\star^2 r_0 + C_\star J \lambda_{N,J} < \frac{1}{2}.$$

For any $\xi, \xi' \in \mathcal{B}$, $\xi' + u(\xi - \xi') \in \mathcal{B}$ for $0 \le u \le 1$. Integrating the Jacobian along u gives

$$\|\Phi(\xi) - \Phi(\xi')\|_\infty = \left\| \int_0^1 \partial_\xi \Phi\big(\xi' + u(\xi - \xi')\big)(\xi - \xi')\, du \right\|_\infty \le \frac{1}{2} \|\xi - \xi'\|_\infty,$$

so Φ is a contraction.

Since $\partial_\xi L_s(x, \xi^*; \theta) = 0$,

$$\|\Phi(\xi^*) - \xi^*\|_\infty = \lambda_{N,J} \|H_s(\xi^*)^{-1} \xi^*\|_\infty \le C_\star^2 J \log J \lambda_{N,J} \le \frac{r_0}{4}.$$

Thus, for every $\xi \in \mathcal{B}$,

$$\|\Phi(\xi) - \xi^*\|_\infty \le \frac{1}{2} \|\xi - \xi^*\|_\infty + \|\Phi(\xi^*) - \xi^*\|_\infty \le \frac{3 r_0}{4} < r_0,$$

so Φ maps $\mathcal{B}$ into itself. Banach's fixed-point theorem gives a unique fixed point $\xi_L \in \mathcal{B}$, hence a unique solution to the FOC. Applying the same inequality at ξ_L and rearranging gives $\|\xi_L - \xi^*\|_\infty \le 2\lambda_{N,J} \|H_s(\xi^*)^{-1} \xi^*\|_\infty$, yielding (6).

The Hessian of $-L_s^{\mathrm{R}}$ is positive definite on $\mathcal{B}$. Therefore, ξ_L is the unique maximizer of L_s^{R} on $\mathcal{B}$.

Since $H_s(\xi) + \lambda_{N,J} I_J = H_s(\xi^*)(I_J - \partial_\xi \Phi(\xi))$ and $\|\partial_\xi \Phi(\xi)\|_\infty < \frac{1}{2}$, the Neumann series

gives

$$(H_s(\xi) + \lambda_{N,J} I_J)^{-1} = \sum_{m=0}^{\infty} (\partial_\xi \Phi(\xi))^m \, H_s(\xi^*)^{-1}.$$

Taking norms yields the inverse bound in (7):

$$\left\|(H_s(\xi) + \lambda_{N,J} I_J)^{-1}\right\|_\infty \le \frac{\|H_s(\xi^*)^{-1}\|_\infty}{1 - \|\partial_\xi \Phi(\xi)\|_\infty} \le 2C_\star J.$$

By Assumption 5(iv), ξ_L is the unique global maximizer. □

Proof of Lemma 1. Since $-L_s^{\mathrm{R}}$ is continuous and coercive, a global maximizer exists. Concavity of every log-share and the quadratic penalty make L_s^{R} strongly concave, so this maximizer is unique for every $s \in \Delta^J$.

For Assumption 5(iv), let $s \in \mathcal{S}_J^+$. Lemma 3 gives ξ_L satisfying

$$\partial_\xi L_s^{\mathrm{R}}(x, \xi_L; \theta, \lambda_{N,J}) = 0, \qquad \|\xi_L - \xi^{\mathrm{BLP}}(x, \theta, s)\|_\infty \le \frac{r_0}{2}.$$

By concavity, ξ_L is the global maximizer. For any ξ with $\|\xi - \xi^{\mathrm{BLP}}(x, \theta, s)\|_\infty \ge r_0$, the triangle inequality gives $\|\xi - \xi_L\|_\infty \ge \frac{r_0}{2}$. Strong concavity therefore implies

$$L_s^{\mathrm{R}}(x, \xi_L; \theta, \lambda_{N,J}) - L_s^{\mathrm{R}}(x, \xi; \theta, \lambda_{N,J}) \ge \frac{\lambda_{N,J}}{2} \|\xi - \xi_L\|_2^2 \ge \frac{\lambda_{N,J} r_0^2}{8}.$$

Since ξ_L lies inside the r_0-ball and the bound holds uniformly over x, s, θ, Assumption 5(iv) follows. □

Lemma 4. *Let $s^0 \in \mathcal{S}_J$ and $Q \sim \mathrm{Multinomial}(N; s^0)$. Define*

$$D_N := J \max_{j \in \mathcal{J}_+} |\hat{s}_j - s_j^0| + |\hat{s}_0 - s_0^0|.$$

There is a constant $C_1 > 0$, depending only on C, such that

$$\mathbb{E}[D_N^2 \mid N, s^0] \le C_1 \left(\frac{J \log J}{N} + \frac{J^2 \log^2 J}{N^2} \right). \tag{8}$$

For every $\varepsilon \in (0, 1]$,

$$\Pr(D_N > \varepsilon \mid N, s^0) \le 2J \exp\left(-\frac{\varepsilon^2 N}{8(C+1)J} \right) + 2 \exp\left(-\frac{\varepsilon^2 N}{2} \right). \tag{9}$$

If $\varepsilon \le \frac{1}{2}\min\{c, c_0\}$, then $D_N \le \varepsilon$ implies $\hat{s} \in \mathcal{S}_J^+$. Moreover,

$$\sup_{s^0 \in \mathcal{S}_J} \Pr(\hat{s} \notin \mathcal{S}_J^+ \mid N, s^0) \le 2J \exp\left(-\frac{cN}{12J}\right) + 2\exp\left(-\frac{c_0 N}{12}\right). \tag{10}$$

Proof. Each $N\hat{s}_j$ is binomial with success probability s_j^0. For $j \in \mathcal{J}_+$, $s_j^0 \le \frac{C}{J}$, so $\mathrm{Var}(\hat{s}_j \mid N, s^0) = \frac{s_j^0(1-s_j^0)}{N} \le \frac{C}{NJ}$. Bernstein's inequality and a union bound over the inside products give, for $u \ge 0$,

$$\Pr\left(J \max_{j \in \mathcal{J}_+} |\hat{s}_j - s_j^0| > \sqrt{\frac{2CJ(\log J + u)}{N}} + \frac{2J(\log J + u)}{3N} \;\middle|\; N, s^0 \right) \le 2\exp(-u).$$

Integrating this bound gives

$$\mathbb{E}\left[\left(J \max_{j \in \mathcal{J}_+} |\hat{s}_j - s_j^0| \right)^2 \;\middle|\; N, s^0 \right] \le \frac{4CJ}{N}(\log J + 2) + \frac{8J^2}{9N^2}(\log J + 2)^2.$$

The RHS is at most $C_1(\frac{J \log J}{N} + \frac{J^2 \log^2 J}{N^2})$. Also, $\mathbb{E}[(\hat{s}_0 - s_0^0)^2 \mid N, s^0] = \frac{s_0^0(1-s_0^0)}{N} \le \frac{1}{4N}$. Combining these bounds and increasing C_1 if necessary gives (8).

If $D_N > \varepsilon$, then $\max_{j \in \mathcal{J}_+} |\hat{s}_j - s_j^0| > \frac{\varepsilon}{2J}$ or $|\hat{s}_0 - s_0^0| > \frac{\varepsilon}{2}$. Bernstein's inequality for the inside shares, Hoeffding's inequality for the outside share, and a union bound give

$$\Pr(D_N > \varepsilon \mid N, s^0) \le 2J \exp\left(-\frac{\varepsilon^2 N}{8J(C + \frac{\varepsilon}{6})}\right) + 2\exp\left(-\frac{\varepsilon^2 N}{2}\right).$$

Since $\varepsilon \le 1$, replacing $C + \frac{\varepsilon}{6}$ by $C + 1$ gives (9).

If $D_N \le \varepsilon \le \frac{1}{2}\min\{c, c_0\}$, then

$$\frac{c}{2J} \le \frac{c - \varepsilon}{J} \le \hat{s}_j \le \frac{C + \varepsilon}{J} \le \frac{2C}{J} \quad (j \in \mathcal{J}_+), \qquad \hat{s}_0 \ge c_0 - \varepsilon \ge \frac{c_0}{2},$$

so $\hat{s} \in \mathcal{S}_J^+$. Finally, the multiplicative Chernoff bound gives

$$\Pr\left(\left| \frac{\hat{s}_j}{s_j^0} - 1 \right| > \frac{1}{2} \;\middle|\; N, s^0 \right) \le 2\exp\left(-\frac{N s_j^0}{12}\right), \qquad j \in \mathcal{J}.$$

If $\left|\frac{\hat{s}_j}{s_j^0} - 1\right| \le \frac{1}{2}$ for every $j \in \mathcal{J}$, then $\hat{s} \in \mathcal{S}_J^+$. A union bound, $s_j^0 \ge \frac{c}{J}$ for inside products, and $s_0^0 \ge c_0$ give (10). □

Proof of Theorem 1. The bounds below hold uniformly over $x \in \mathcal{X}^J$, $\theta \in \Theta$, and $s^0 \in \mathcal{S}_J$.

Let D_N be defined in Lemma 4 and fix $\varepsilon = \min\{\frac{c}{2}, \frac{c_0}{2}, \frac{r_0}{2C_\star^2}\}$. Choose C_1 sufficiently large and $0 < c_1 \leq \min\{\frac{1}{2C_\star}, \frac{\varepsilon^2}{8(C+1)}, \frac{\varepsilon^2}{2}\}$. Then, it holds that (8), the tail bound $\Pr(D_N > \varepsilon \mid N, s^0) \leq C_1(J\exp(-\frac{c_1 N}{J}) + \exp(-c_1 N))$, and $\hat{s} \in \mathcal{S}_J^+$ whenever $D_N \leq \varepsilon$.

On this event, Lemma 2 gives $\|\xi^{\mathrm{BLP}}(x,\theta,\hat{s}) - \xi^{\mathrm{BLP}}(x,\theta,s^0)\|_\infty \leq C_\star^2 D_N \leq \frac{r_0}{2}$. Lemma 3 and regularized separation give a unique global maximizer on this event. It equals $\xi_N^{\mathrm{R}}(x,\theta,\hat{s})$ for every selection and satisfies (6).

For $0 \leq u \leq 1$, define $\xi(u) := (1-u)\xi^{\mathrm{BLP}}(x,\theta,s^0) + u\xi_N^{\mathrm{R}}(x,\theta,\hat{s})$. The preceding bound and (6) imply $\|\xi(u) - \xi^{\mathrm{BLP}}(x,\theta,\hat{s})\|_\infty \leq (1-u)\frac{r_0}{2} + u\frac{r_0}{2} = \frac{r_0}{2}$. Thus the curvature bound in (7) applies at every $\xi(u)$. Define $\overline{H} := \int_0^1 H_{\hat{s}}(x,\xi(u);\theta)\,du \succeq \frac{1}{2C_\star J} I_J$. Consequently,

$$\|(\overline{H} + \lambda_{N,J} I_J)^{-1}\|_2 \leq \frac{J}{\frac{1}{2C_\star} + J\lambda_{N,J}} \leq \frac{J}{c_1 + J\lambda_{N,J}}.$$

The regularized FOC and the mean-value theorem give

$$(\overline{H} + \lambda_{N,J} I_J)(\xi_N^{\mathrm{R}} - \xi^{\mathrm{BLP}}(s^0)) = \partial_\xi L_{\hat{s}}(\xi^{\mathrm{BLP}}(s^0)) - \lambda_{N,J}\xi^{\mathrm{BLP}}(s^0),$$

where we suppress (x,θ). Assumptions 5(iii) and 5(i) imply, respectively,

$$\|\partial_\xi L_{\hat{s}}(\xi^{\mathrm{BLP}}(s^0))\|_2 \leq \frac{C_\star}{\sqrt{J}} D_N, \qquad \|\xi^{\mathrm{BLP}}(s^0)\|_2 \leq C_\star\sqrt{J}\log J.$$

Consequently, on $\{D_N \leq \varepsilon\}$,

$$\frac{1}{J\log J}\|\xi_N^{\mathrm{R}} - \xi^{\mathrm{BLP}}(s^0)\|_2^2 \leq \frac{2C_\star^2\left(\frac{D_N^2}{\log J} + J^2\log J\lambda_{N,J}^2\right)}{(c_1 + J\lambda_{N,J})^2}.$$

Taking expectations and using (8), with C_1 increased if needed, yields the first term.

It remains to control $\{D_N > \varepsilon\}$. Fix any $\widetilde{\xi} \in \mathcal{A}_N(x,\theta,\hat{s})$. Optimality of $\widetilde{\xi}$ and $L_{\hat{s}}(x,\widetilde{\xi};\theta) \leq 0$ imply

$$\|\widetilde{\xi}\|_2^2 \leq \|\xi^{\mathrm{BLP}}(x,\theta,s^0)\|_2^2 - \frac{2}{\lambda_{N,J}}\sum_{j\in\mathcal{J}} \hat{s}_j \log s_j^0.$$

Since $\sum_{j\in\mathcal{J}} \hat{s}_j = 1$ and $s^0 \in \mathcal{S}_J$,

$$-\sum_{j\in\mathcal{J}} \hat{s}_j \log s_j^0 \leq \max\left\{\log\left(\frac{J}{c}\right), \log\left(\frac{1}{c_0}\right)\right\} \leq C_1 \log J.$$

The preceding inequalities and Assumption 5(i) give, after increasing C_1,

$$\sup_{x,\theta,\widetilde{\xi}\in\mathcal{A}_N(x,\theta,\hat{s})} \frac{1}{J\log J}\|\widetilde{\xi}-\xi^{\mathrm{BLP}}(x,\theta,s^0)\|_2^2 \le C_1\left(\log J+\frac{1}{J\lambda_{N,J}}\right).$$

Combining this bound with (9) gives the second term in the theorem. □

Proof of Corollary 1. By Lemma 4, $\sup_{s^0\in\mathcal{S}_J}\Pr(\hat{s}\notin\mathcal{S}_J^+ \mid N,s^0)\to 0$ under $\frac{J\log J}{N}\to 0$. On $\{\hat{s}\in\mathcal{S}_J^+\}$, Lemma 3 gives a unique global maximizer simultaneously for all $x\in\mathcal{X}^J$ and $\theta\in\Theta$.

Equation (8) and Markov's inequality give $D_N = O_p(\sqrt{\frac{J\log J}{N}}+\frac{J\log J}{N})$ uniformly over $s^0\in\mathcal{S}_J$. Combining Lemma 2 with (6) on this event gives, uniformly over $s^0\in\mathcal{S}_J$,

$$\sup_{x\in\mathcal{X}^J,\theta\in\Theta}\|\xi_N^{\mathrm{R}}(x,\theta,\hat{s})-\xi^{\mathrm{BLP}}(x,\theta,s^0)\|_\infty = O_p\left(\sqrt{\frac{J\log J}{N}}+\frac{J\log J}{N}+J\log J\lambda_{N,J}\right).$$

□

A.2 Proofs in Section 3

Lemma 5. *Under Assumptions 8(i), 8(iii)–8(iv), and 9(i),*

$$\Pr(\mathcal{E}_T^c)\le \Pr\big(s_t^0\notin\mathcal{S}_{J_t}\text{ for some }t\le T\big)+2TJ_{\max,T}\exp\left(-\frac{cN_{\min,T}}{12J_{\max,T}}\right)+2T\exp\left(-\frac{c_0N_{\min,T}}{12}\right)\longrightarrow 0.$$

Proof. For $s_t^0\in\mathcal{S}_{J_t}$, (10) in Lemma 4 bounds $\Pr(\hat{s}_t\notin\mathcal{S}_{J_t}^+\mid s_t^0)$. Taking expectations and a union bound over markets gives the result. □

Lemma 6. *Suppose Assumptions 8–10 hold. For all sufficiently large T, on $\mathcal{E}_T$, uniformly over $t\le T$ and $\theta\in\Theta$,*

$$\|\xi_{N_t}^{\mathrm{R}}(x_t,\theta,\hat{s}_t)-\xi^{\mathrm{BLP}}(x_t,\theta,\hat{s}_t)\|_\infty\le 2C_\star^2J_t\log J_t\lambda_{N_t,J_t}.$$

There is a constant $C_0\ge 2C_\star^2$, depending only on the constants in the assumptions, such that, uniformly over $t\le T$ and $\theta\in\mathcal{N}_0$,

$$\max_{j\in\mathcal{J}_{t,+}}\left\|\partial_\theta\xi_{j,N_t}^{\mathrm{R}}(x_t,\theta,\hat{s}_t)-\partial_\theta\xi_j^{\mathrm{BLP}}(x_t,\theta,\hat{s}_t)\right\|_2\le C_0J_t(1+\log J_t)\lambda_{N_t,J_t}.$$

Proof. We suppress the arguments $(x_t,\theta,\hat{s}_t)$ of $\xi_{N_t}^{\mathrm{R}}$ and ξ^{BLP}. Note that $\max_{t\le T}J_t(1+$

$\log J_t)\lambda_{N_t,J_t} = o(1)$. On $\mathcal{E}_T$, (6) gives the first bound.

Write $H^{\mathrm{R}}_{\hat{s}_t}, C^{\mathrm{R}}_{\hat{s}_t}$ for the negative Hessian and mixed derivative of $L^{\mathrm{R}}_{\hat{s}_t}$ at $\xi^{\mathrm{R}}_{N_t}$, and $H_{\hat{s}_t}, C_{\hat{s}_t}$ for those of $L_{\hat{s}_t}$ at ξ^{BLP}, with $\theta \in \mathcal{N}_0$. Equation (7) gives $\|(H^{\mathrm{R}}_{\hat{s}_t})^{-1}\|_\infty \le 2C_\star J_t$. Implicit differentiation gives

$$\partial_\theta \xi^{\mathrm{R}}_{N_t} = (H^{\mathrm{R}}_{\hat{s}_t})^{-1} C^{\mathrm{R}}_{\hat{s}_t}, \qquad \partial_\theta \xi^{\mathrm{BLP}} = H^{-1}_{\hat{s}_t} C_{\hat{s}_t}.$$

Note that

$$\partial_\theta \xi^{\mathrm{R}}_{N_t} - \partial_\theta \xi^{\mathrm{BLP}} = (H^{\mathrm{R}}_{\hat{s}_t})^{-1} \left(C^{\mathrm{R}}_{\hat{s}_t} - C_{\hat{s}_t} + (H_{\hat{s}_t} - H^{\mathrm{R}}_{\hat{s}_t}) H^{-1}_{\hat{s}_t} C_{\hat{s}_t} \right).$$

For matrices with compatible dimensions, $\max_j \|[AB]_{j\cdot}\|_2 \le \|A\|_\infty \max_j \|[B]_{j\cdot}\|_2$. In particular, Assumptions 5(ii) and 9(ii) give $\max_j \|[H^{-1}_{\hat{s}_t} C_{\hat{s}_t}]_{j\cdot}\|_2 \le C^2_\star$. Applying it, Assumption 5(ii), and Assumption 9(ii) to the equation above gives

$$2(C^2_\star + C^4_\star)\|\xi^{\mathrm{R}}_{N_t} - \xi^{\mathrm{BLP}}\|_\infty + 2C^3_\star J_t \lambda_{N_t,J_t} \le C_0 J_t (1 + \log J_t)\lambda_{N_t,J_t}.$$

□

Lemma 7. *Suppose Assumptions 8–9 hold. Under Assumption 10(i),*

$$\sup_{\theta\in\Theta} \|\widehat{G}^{\mathrm{R}}_T(\theta) - G^{\mathrm{BLP}}_T(\theta)\|_2 + \sup_{\theta\in\mathcal{N}_0} \|\partial_\theta \widehat{G}^{\mathrm{R}}_T(\theta) - \partial_\theta G^{\mathrm{BLP}}_T(\theta)\|_2 = O_p\left(\frac{J_{\max,T} \log J_{\max,T}}{N_{\min,T}} \right).$$

Under Assumption 10(ii), the right-hand side (RHS) improves to $O_p(\frac{J_{\max,T}}{N_{\min,T}})$.

Proof. Lemma 5 gives $\Pr(\mathcal{E}_T) \to 1$. Let $\overline{Z}_T := \frac{1}{M_T}\sum_{t=1}^T \sum_{j\in\mathcal{J}_{t,+}} \|z_{jt}\|_2$. Assumption 8(ii), Jensen's inequality, and Markov's inequality give $\overline{Z}_T = O_p(1)$. Lemma 6 therefore implies

$$\sup_{\theta\in\Theta} \|\widehat{G}^{\mathrm{R}}_T(\theta) - G^{\mathrm{BLP}}_T(\theta)\|_2 + \sup_{\theta\in\mathcal{N}_0} \|\partial_\theta \widehat{G}^{\mathrm{R}}_T(\theta) - \partial_\theta G^{\mathrm{BLP}}_T(\theta)\|_2 \le 2C_0 \overline{Z}_T \max_{t\le T} J_t(1 + \log J_t)\lambda_{N_t,J_t}.$$

Under Assumption 10(i), bound is $O(\frac{J_{\max,T}\log J_{\max,T}}{N_{\min,T}})$. Under Assumption 10(ii), the bound is $O\left(\frac{J_{\max,T}}{N_{\min,T}} + \frac{J^2_{\max,T}\log J_{\max,T}}{N^2_{\min,T}}\right) = O\left(\frac{J_{\max,T}}{N_{\min,T}}\right)$. □

Proof of Theorem 2. Continuity of Q and compactness of Θ give $\sup_{\theta\in\Theta}|Q(\theta)| < \infty$. Assumptions 11(i)–11(ii) therefore imply $\sup_{\theta\in\Theta}\|G^{\mathrm{BLP}}_T(\theta)\|_2 = O_p(1)$. Lemma 7 gives $\sup_{\theta\in\Theta}\|\widehat{G}^{\mathrm{R}}_T(\theta) - G^{\mathrm{BLP}}_T(\theta)\|_2 = o_p(1)$. Hence

$$\begin{aligned}
&\sup_{\theta\in\Theta} \left| \widehat{G}^{\mathrm{R}}_T(\theta)' W_T \widehat{G}^{\mathrm{R}}_T(\theta) - G^{\mathrm{BLP}}_T(\theta)' W_T G^{\mathrm{BLP}}_T(\theta) \right| \\
&\quad \le \|W_T\|_2 \sup_{\theta\in\Theta} \|\widehat{G}^{\mathrm{R}}_T(\theta) - G^{\mathrm{BLP}}_T(\theta)\|_2 \left(2\sup_{\theta\in\Theta}\|G^{\mathrm{BLP}}_T(\theta)\|_2 + \sup_{\theta\in\Theta}\|\widehat{G}^{\mathrm{R}}_T(\theta) - G^{\mathrm{BLP}}_T(\theta)\|_2 \right) = o_p(1).
\end{aligned}$$

Together with Assumption 11(ii), this gives $\sup_{\theta\in\Theta} |\widehat{G}_T^{\mathrm{R}}(\theta)' W_T \widehat{G}_T^{\mathrm{R}}(\theta) - Q(\theta)| = o_p(1)$. Since Θ is compact and Q is continuous with a unique minimizer at θ_0, Theorem 2.1 of Newey and McFadden (1994) gives $\widehat{\theta}_T^{\mathrm{R}} \xrightarrow{p} \theta_0$. □

Proof of Theorem 3. Lemma 7 gives

$$\sqrt{T} \sup_{\theta\in\mathcal{N}_0} \left(\|\widehat{G}_T^{\mathrm{R}}(\theta) - G_T^{\mathrm{BLP}}(\theta)\|_2 + \|\partial_\theta \widehat{G}_T^{\mathrm{R}}(\theta) - \partial_\theta G_T^{\mathrm{BLP}}(\theta)\|_2 \right) = o_p(1).$$

The moment CLT in Assumption 11(iv) implies $\sqrt{T}\, G_T^{\mathrm{BLP}}(\theta_0) = O_p(1)$. Together with the preceding approximation and the Jacobian condition in the same assumption, this gives uniform convergence of the regularized Jacobian to Γ on $\mathcal{N}_0$ and $\sqrt{T}\, \widehat{G}_T^{\mathrm{R}}(\theta_0) = O_p(1)$.

By Theorem 2, $\widehat{\theta}_T^{\mathrm{R}} \in \mathcal{N}_0$ with probability approaching one. Since $\mathcal{N}_0$ is convex, $\theta_0 + u(\widehat{\theta}_T^{\mathrm{R}} - \theta_0) \in \mathcal{N}_0$ for every $u \in [0,1]$ on this event. Applying the chain rule and integrating with respect to u gives

$$\widehat{G}_T^{\mathrm{R}}(\widehat{\theta}_T^{\mathrm{R}}) = \widehat{G}_T^{\mathrm{R}}(\theta_0) + \left[\int_0^1 \partial_\theta \widehat{G}_T^{\mathrm{R}} \left(\theta_0 + u(\widehat{\theta}_T^{\mathrm{R}} - \theta_0) \right) du \right] (\widehat{\theta}_T^{\mathrm{R}} - \theta_0).$$

Every point along this path is within distance $\|\widehat{\theta}_T^{\mathrm{R}} - \theta_0\|_2 = o_p(1)$ of θ_0. Uniform Jacobian convergence and continuity of Γ therefore give

$$\sup_{0\le u\le 1} \left\| \partial_\theta \widehat{G}_T^{\mathrm{R}} \left(\theta_0 + u(\widehat{\theta}_T^{\mathrm{R}} - \theta_0) \right) - \Gamma_0 \right\|_2 = o_p(1).$$

Integrating this bound shows that the integral converges in probability to Γ_0. Taking $u = 1$ gives the same limit for $\partial_\theta \widehat{G}_T^{\mathrm{R}}(\widehat{\theta}_T^{\mathrm{R}})$.

At the interior minimizer $\widehat{\theta}_T^{\mathrm{R}}$, the GMM FOC is $\partial_\theta \widehat{G}_T^{\mathrm{R}}(\widehat{\theta}_T^{\mathrm{R}})' W_T \widehat{G}_T^{\mathrm{R}}(\widehat{\theta}_T^{\mathrm{R}}) = 0$. Substituting the integral expansion gives

$$\left[\partial_\theta \widehat{G}_T^{\mathrm{R}}(\widehat{\theta}_T^{\mathrm{R}})' W_T \int_0^1 \partial_\theta \widehat{G}_T^{\mathrm{R}} \left(\theta_0 + u(\widehat{\theta}_T^{\mathrm{R}} - \theta_0) \right) du \right] (\widehat{\theta}_T^{\mathrm{R}} - \theta_0) = -\partial_\theta \widehat{G}_T^{\mathrm{R}}(\widehat{\theta}_T^{\mathrm{R}})' W_T \widehat{G}_T^{\mathrm{R}}(\theta_0).$$

The matrix in square brackets converges in probability to $\Gamma_0' W \Gamma_0$, which is positive definite by Assumptions 11(i) and 11(iv). Its inverse therefore exists with probability approaching one and converges to $(\Gamma_0' W \Gamma_0)^{-1}$. Multiplying by this inverse and by $\sqrt{T}$, and using $\sqrt{T}\, \widehat{G}_T^{\mathrm{R}}(\theta_0) = O_p(1)$, gives

$$\sqrt{T}(\widehat{\theta}_T^{\mathrm{R}} - \theta_0) = -(\Gamma_0' W \Gamma_0)^{-1} \Gamma_0' W \sqrt{T}\, \widehat{G}_T^{\mathrm{R}}(\theta_0) + o_p(1).$$

Assumptions 11(i)–11(iii) and Theorem 2.1 of Newey and McFadden (1994) give $\widehat{\theta}_T^{\mathrm{BLP}} \xrightarrow{p} \theta_0$.

The same expansion therefore holds for G_T^{BLP} and $\widehat{\theta}_T^{\mathrm{BLP}}$.

Subtracting the two expansions and using Lemma 7 at θ_0 gives $\sqrt{T}(\widehat{\theta}_T^{\mathrm{R}} - \widehat{\theta}_T^{\mathrm{BLP}}) = o_p(1)$. The moment CLT in Assumption 11(iv) and the BLP expansion give $\sqrt{T}(\widehat{\theta}_T^{\mathrm{BLP}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_{\mathrm{BLP}})$ by Slutsky's theorem. First-order equivalence then gives the same limit for $\widehat{\theta}_T^{\mathrm{R}}$. □

B Logit Demand Systems

We consider the following three logit demand systems. The utility adjustments $-\log J$ for logit and $-(1-\rho)\log J$ for nested logit correspond to the full-congestion specification of Ackerberg and Rysman (2005, pp. 779–780). Let $\nu \in \mathrm{supp}(F) \subseteq \mathbb{R}^R$ denote consumer tastes. We suppress arguments other than ν in individual probabilities when there is no ambiguity.

Simple logit. The shares are

$$\sigma_j(x,\xi;\beta) = \frac{\exp(x_j'\beta + \xi_j - \log J)}{1 + \sum_{k\in\mathcal{J}_+}\exp(x_k'\beta + \xi_k - \log J)}, \quad \sigma_0(x,\xi;\beta) = \frac{1}{1 + \sum_{k\in\mathcal{J}_+}\exp(x_k'\beta + \xi_k - \log J)}.$$

Random-coefficients logit. Normalize ν so that $\mathbb{E}\nu = 0$ and $\mathbb{E}[\nu\nu'] = I_R$. Individual inside utility is $x_j'\beta + x_j'\Sigma\nu + \xi_j - \log J + \varepsilon_{ij}$, and outside utility is ε_{i0}. The individual probabilities $\sigma_j(\nu; x, \xi, \theta)$ are given by the simple-logit formulas with $x_j'\beta$ replaced by $x_j'\beta + x_j'\Sigma\nu$, and aggregate shares are $\sigma_j(x,\xi;\theta) = \int \sigma_j(\nu; x, \xi, \theta)dF(\nu)$.

Random-coefficients nested logit. Place all inside goods in a purchase nest and keep the outside option separate. Let $\rho \in [0, \rho_{\max}]$, with $\rho_{\max} < 1$, be the nesting parameter. With $\delta_j(\nu) = x_j'\beta + x_j'\Sigma\nu + \xi_j - (1-\rho)\log J$, define

$$D(\nu) = \sum_{k\in\mathcal{J}_+}\exp\left(\frac{\delta_k(\nu)}{1-\rho}\right), \quad \sigma_{j|1}(\nu) = \frac{\exp\left(\frac{\delta_j(\nu)}{1-\rho}\right)}{D(\nu)}, \quad \sigma_{\mathrm{buy}}(\nu) = \frac{D(\nu)^{1-\rho}}{1 + D(\nu)^{1-\rho}}.$$

The individual probabilities are $\sigma_j(\nu) = \sigma_{j|1}(\nu)\sigma_{\mathrm{buy}}(\nu)$ and $\sigma_0(\nu) = 1 - \sigma_{\mathrm{buy}}(\nu)$, and aggregate shares are their integrals.

B.1 Common results

The results below include simple and RC logit by setting $\rho = 0$. They also apply to finite mixtures with type-specific nesting parameters, since the bounds depend only on $\rho_{\max}$. In the proofs below, $\widetilde{C} \geq 1$ denotes finite upper-bound constants that may increase between inequalities and depend only on the share bounds and model restrictions. We write $D_\xi f(\xi)[w] := \frac{d}{du} f(\xi + uw)\big|_{u=0}$ for the directional derivative of f in direction $w \in \mathbb{R}^J$, holding other arguments fixed.

Lemma 8. *For each of the three models, suppose that, at $\xi^{\mathrm{BLP}}(x, \theta, s)$,*

$$\frac{\int \sigma_j(\nu)\sigma_0(\nu)dF(\nu)}{s_j} \geq \eta > 0, \qquad j \in \mathcal{J}_+,$$

uniformly over $x \in \mathcal{X}^J$, $s \in \mathcal{S}_J^+$, and $\theta \in \Theta$. Then the inverse-information bound in Assumption 5(ii) holds.

Proof. Let D_s be the $J \times J$ inside-share Jacobian at $\xi^{\mathrm{BLP}}(x, \theta, s)$ and let $s_+ = (s_1, \ldots, s_J)'$. Differentiation of individual nested-logit probabilities gives

$$\partial_{\xi_k}\sigma_j(\nu) = \sigma_j(\nu)\left(\frac{\mathbf{1}(j=k)}{1-\rho} - \left(\frac{\rho}{1-\rho} + \sigma_{\mathrm{buy}}(\nu)\right)\sigma_{k|1}(\nu)\right), \ \partial_{\xi_k}\sigma_0(\nu) = -\sigma_0(\nu)\sigma_k(\nu). \quad (11)$$

The inside Jacobian is symmetric, its off-diagonal entries are nonpositive, and its jth row sum is $\int \sigma_j(\nu)\sigma_0(\nu)dF(\nu)$. Since $s \in \mathcal{S}_J^+$, its diagonal-dominance margin satisfies

$$(D_s)_{jj} - \sum_{k \neq j} |(D_s)_{jk}| = \int \sigma_j(\nu)\sigma_0(\nu)dF(\nu) \geq \eta s_j \geq \frac{\eta c}{2J}.$$

For any $w \in \mathbb{R}^J$, choose j such that $|w_j| = \|w\|_\infty$. Then

$$\|D_s w\|_\infty \geq \left((D_s)_{jj} - \sum_{k \neq j} |(D_s)_{jk}|\right)\|w\|_\infty \geq \frac{\eta c}{2J}\|w\|_\infty.$$

Hence D_s is nonsingular and $\|D_s^{-1}\|_\infty \leq \frac{2J}{\eta c}$.

Define $\Omega(s) = \mathrm{diag}(s_+) - s_+ s_+'$. Equation (5) gives

$$H_s = D_s'\Omega(s)^{-1}D_s, \qquad \Omega(s)^{-1} = \mathrm{diag}(s_+)^{-1} + \frac{1}{s_0}\mathbf{1}_J\mathbf{1}_J'.$$

Since D_s is symmetric and $\|\Omega(s)\|_\infty \le 2\max_{j\in\mathcal{J}_+} s_j \le \frac{4C}{J}$, we have

$$\|H_s^{-1}\|_\infty = \|D_s^{-1}\Omega(s)D_s^{-1}\|_\infty \le \frac{16C}{\eta^2 c^2}J.$$

□

Lemma 9. *For each of the three models, the Hessian-Lipschitz condition in Assumption 5(ii) and Assumption 5(iii) hold. For nested logit the constants may depend on $\rho_{\max}$.*

Proof. Take $0 < r_0 \le 1$. Equation (11) and $\sum_{k\in\mathcal{J}_+}\sigma_{k|1}(\nu) = 1$ give

$$\sum_{k\in\mathcal{J}_+} |\partial_{\xi_k}\log\sigma_j(\nu)| \le \frac{2}{1-\rho_{\max}}, \qquad j\in\mathcal{J}.$$

For $\|\xi - \xi^{\mathrm{BLP}}(x,\theta,s)\|_\infty \le r_0$, let $\xi(u) = \xi^{\mathrm{BLP}}(x,\theta,s) + u(\xi - \xi^{\mathrm{BLP}}(x,\theta,s))$, $0\le u\le 1$. Integrating along this path gives

$$\left|\log\frac{\sigma_j(\nu;x,\xi,\theta)}{\sigma_j(\nu;x,\xi^{\mathrm{BLP}}(x,\theta,s),\theta)}\right| \le \int_0^1 \sum_{k\in\mathcal{J}_+} |\partial_{\xi_k}\log\sigma_j(\nu;x,\xi(u),\theta)||\xi_k - \xi_k^{\mathrm{BLP}}(x,\theta,s)|\,du \le \frac{2r_0}{1-\rho_{\max}}.$$

Exponentiating and integrating against F, using the definition of $\xi^{\mathrm{BLP}}(x,\theta,s)$, yields

$$\widetilde{C}^{-1}s_j \le \sigma_j(x,\xi;\theta) \le \widetilde{C}s_j, \qquad j\in\mathcal{J}. \tag{12}$$

Write $h_{jk}(\nu) = \partial_{\xi_k}\log\sigma_j(\nu)$, obtained from (11). For $\|w\|_\infty \le 1$,

$$D_\xi\sigma_{k|1}[w] = \frac{\sigma_{k|1}}{1-\rho}\left(w_k - \sum_{\ell\in\mathcal{J}_+}\sigma_{\ell|1}w_\ell\right), \qquad D_\xi\sigma_{\mathrm{buy}}[w] = \sigma_{\mathrm{buy}}\sigma_0\sum_{\ell\in\mathcal{J}_+}\sigma_{\ell|1}w_\ell.$$

The product rule, together with $\sum_{\ell\in\mathcal{J}_+}|\partial_{\xi_\ell}\sigma_{k|1}| \le \frac{2\sigma_{k|1}}{1-\rho}$ and $\sum_{\ell\in\mathcal{J}_+}|\partial_{\xi_\ell}\sigma_{\mathrm{buy}}| \le \sigma_{\mathrm{buy}}$, gives

$$|h_{jk}| + \sum_{\ell\in\mathcal{J}_+}(|\partial_{\xi_\ell}h_{jk}| + |D_\xi\partial_{\xi_\ell}h_{jk}[w]|) \le \widetilde{C}(\mathbf{1}(j=k) + \sigma_{k|1}), \qquad j\in\mathcal{J}_+.$$

For $j = 0$, the RHS is $\widetilde{C}\sigma_k$.

Let $\mathbb{E}_j$ denote integration against $\frac{\sigma_j(\nu)}{\sigma_j(x,\xi;\theta)}dF(\nu)$. Equation (12) implies

$$\sum_{j\in\mathcal{J}_+} s_j\mathbb{E}_j\sigma_{k|1} \le \widetilde{C}\int\sigma_{\mathrm{buy}}(\nu)\sigma_{k|1}(\nu)dF(\nu) = \widetilde{C}\sigma_k(x,\xi;\theta) \le \frac{2C\widetilde{C}^2}{J}, \quad s_0\mathbb{E}_0\sigma_k \le \frac{2C\widetilde{C}^2}{J}.$$

Hessian-Lipschitz bound. Differentiation gives

$$D_\xi \mathbb{E}_j f[w] = \mathbb{E}_j D_\xi f[w] + \mathrm{Cov}_j(f, D_\xi \log \sigma_j[w]). \tag{13}$$

In particular,

$$\partial^2_{\xi_\ell \xi_k} \log \sigma_j(x, \xi; \theta) = \mathbb{E}_j \partial_{\xi_\ell} h_{jk} + \mathrm{Cov}_j(h_{jk}, h_{j\ell}).$$

Summing over ℓ and differentiating once more gives, for $j \in \mathcal{J}_+$,

$$\sum_{\ell \in \mathcal{J}_+} \left(|\partial^2_{\xi_\ell \xi_k} \log \sigma_j(x, \xi; \theta)| + |D_\xi \partial^2_{\xi_\ell \xi_k} \log \sigma_j(x, \xi; \theta)[w]| \right) \le \widetilde{C} \mathbb{E}_j[\mathbf{1}(j = k) + \sigma_{k|1}].$$

For $j = 0$, the RHS is $\widetilde{C} \mathbb{E}_0 \sigma_k$. Since $H_s = -\sum_{j \in \mathcal{J}} s_j \partial^2_{\xi \xi'} \log \sigma_j$, the triangle inequality and the preceding bounds give

$$\|D_\xi H_s[w]\|_\infty \le \widetilde{C} \max_{k \in \mathcal{J}_+} \left(s_k + \sum_{j \in \mathcal{J}_+} s_j \mathbb{E}_j \sigma_{k|1} + s_0 \mathbb{E}_0 \sigma_k \right) \le \frac{2C\widetilde{C} + 4C\widetilde{C}^3}{J}.$$

For ξ, ξ' in the same ball, $\xi' + u(\xi - \xi')$ remains in that ball for $0 \le u \le 1$. Note that

$$H_s(x, \xi; \theta) - H_s(x, \xi'; \theta) = \int_0^1 D_\xi H_s(x, \xi' + u(\xi - \xi'); \theta)[\xi - \xi'] \, du.$$

For $\xi \neq \xi'$, take $w = \frac{\xi - \xi'}{\|\xi - \xi'\|_\infty}$. Then $\|w\|_\infty = 1$ and $D_\xi H_s[\xi - \xi'] = \|\xi - \xi'\|_\infty D_\xi H_s[w]$. Taking norms in the integral yields

$$\begin{aligned} \|H_s(x, \xi; \theta) - H_s(x, \xi'; \theta)\|_\infty &\le \|\xi - \xi'\|_\infty \int_0^1 \|D_\xi H_s(x, \xi' + u(\xi - \xi'); \theta)[w]\|_\infty \, du \\ &\le \frac{2C\widetilde{C} + 4C\widetilde{C}^3}{J} \|\xi - \xi'\|_\infty. \end{aligned}$$

Score bounds. For every $k \in \mathcal{J}_+$, differentiating the aggregate share gives

$$\partial_{\xi_k} \log \sigma_j(x, \xi; \theta) = \frac{\int \sigma_j(\nu) h_{jk}(\nu) dF(\nu)}{\sigma_j(x, \xi; \theta)} = \mathbb{E}_j h_{jk}.$$

For inside products, $|h_{jk}| \le \widetilde{C}(\mathbf{1}(j = k) + \sigma_{k|1})$, and the indicator sums to one. Since $s_j \ge \frac{c}{2J}$,

the preceding weighted bound gives

$$\sum_{j\in\mathcal{J}_+}|\partial_{\xi_k}\log\sigma_j(x,\xi;\theta)|\le\widetilde{C}\left(1+\sum_{j\in\mathcal{J}_+}\mathbb{E}_j\sigma_{k|1}\right)\le\widetilde{C}\left(1+\frac{2J}{c}\sum_{j\in\mathcal{J}_+}s_j\mathbb{E}_j\sigma_{k|1}\right)\le\widetilde{C}\left(1+\frac{4C\widetilde{C}^2}{c}\right).$$

For the outside product, $h_{0k}=-\sigma_k$ and $s_0\ge\frac{c_0}{2}$ give

$$|\partial_{\xi_k}\log\sigma_0(x,\xi;\theta)|=\mathbb{E}_0\sigma_k\le\frac{2}{c_0}s_0\mathbb{E}_0\sigma_k\le\frac{4C\widetilde{C}^2}{c_0J}.$$

□

B.2 Model-specific results

For each model below, choose a common $C_\star\ge1$ to dominate the derived inverse, derivative, and global recovery bounds, then choose $0<r_0<\frac{1}{4C_\star^2}$.

B.2.1 Simple logit

Proposition 2. *Simple logit satisfies Assumptions 1–3. Under Assumptions 4 and 6, it satisfies Assumption 5, with constants uniform in J.*

Proof. Assumptions 1–3 follow directly from the simple-logit formulas and the explicit inverse

$$\xi_j^{\mathrm{BLP}}(x,\beta,s)=\log\left(\frac{Js_j}{s_0}\right)-x_j'\beta.$$

Under Assumption 4, this inverse is uniformly bounded for $s\in\mathcal{S}_J^+$. Since the outside-overlap ratio is $s_0\ge\frac{c_0}{2}$, Lemmas 8 and 9 give Assumptions 5(ii)–5(iii). Log-concavity of the simple-logit shares and Assumption 6 imply Assumption 5(iv) by Lemma 1. □

B.2.2 Random-coefficients logit and nested logit

We consider two cases: continuously distributed tastes and a finite mixture of consumer types.

In the continuous case, the coefficient vector is $\beta+\Sigma\nu$, where $\beta\in\mathbb{R}^K$ and $\Sigma\in\mathbb{R}^{K\times R}$ are parameters. The distribution F is fixed. Thus β is the mean coefficient vector and Σ determines its covariance.

Fix $c_x \in (0,1]$ and $\alpha \geq 0$. Define

$$\mathcal{D}_J := \left\{ x \in \mathcal{X}^J : \min_{j \in \mathcal{J}_+} \frac{1}{J} \sum_{k \in \mathcal{J}_+} \mathbf{1}\left(\|x_k - x_j\|_2 \leq h \sup_{u,v \in \mathcal{X}} \|u - v\|_2 \right) \geq c_x h^\alpha, \ \forall h \in (0,1] \right\}.$$

At each fixed relative radius h, every product has a positive fraction of products with similar characteristics.

(RC1) *Log-concave tastes.* The distribution F has a log-concave density on a compact convex support, or $\nu \sim \mathcal{N}(0, I_R)$.

(RC2) *Lognormal tastes.* Let $Y_1, \ldots, Y_R$ be independent lognormal variables with fixed distributions and positive variances, and set $\nu_k = \frac{Y_k - \mathbb{E}Y_k}{\sqrt{\mathrm{Var}(Y_k)}}$. We assume $x \in \mathcal{D}_J$, and the results below hold uniformly over $x \in \mathcal{D}_J$.

(RC3) *Finite mixture.* There are M types, with probabilities $\pi_m \geq 0$ satisfying $\sum_{m=1}^M \pi_m = 1$. Type m has coefficient vector $\beta_m \in \mathbb{R}^K$ and nesting parameter $\rho_m \in [0, \rho_{\max}]$. Write $\sigma_j^{(m)}(x, \xi; \beta_m, \rho_m)$ for its nested-logit probability with utility index $x_j'\beta_m + \xi_j - (1 - \rho_m)\log J$. Aggregate shares are

$$\sigma_j(x, \xi; \theta) = \sum_{m=1}^M \pi_m \sigma_j^{(m)}(x, \xi; \beta_m, \rho_m), \qquad j \in \mathcal{J}.$$

For RC logit, set $\rho_m = 0$ for every type. For finite mixture models, the type probabilities π_m, type-specific coefficients β_m and nesting parameters ρ_m are parameters.

Lemma 10. *Suppose Assumption 4 holds and $\mathbb{E}(1 + \|\nu\|_2)^{4+\alpha} < \infty$. For RC logit and RC nested logit, for every $\varepsilon \in (0, \frac{1}{2})$, there is $C_\varepsilon \geq 1$ such that, uniformly over J, $x \in \mathcal{X}^J$, $\theta \in \Theta$, and ξ with $\varepsilon \leq \sigma_0(x, \xi; \theta) \leq 1 - \varepsilon$,*

$$\sigma_j(x, \xi; \theta) \geq \int \sigma_j(\nu) \sigma_0(\nu) dF(\nu) \geq \frac{\exp\left(\frac{\xi_j}{1-\rho}\right)}{C_\varepsilon J}, \qquad j \in \mathcal{J}_+. \tag{14}$$

If also $x \in \mathcal{D}_J$ and $\min_{j \in \mathcal{J}_+} J\sigma_j(x, \xi; \theta) \geq \varepsilon$, then

$$\max_{j \in \mathcal{J}_+} \frac{\int (1 + \|\nu\|_2)^4 \sigma_j(\nu; x, \xi, \theta) dF(\nu)}{\sigma_j(x, \xi; \theta)} \leq C_\varepsilon, \tag{15}$$

and, for $j \in \mathcal{J}_+$,

$$\frac{\exp\left(\frac{\xi_j}{1-\rho}\right)}{C_\varepsilon J} \leq \sigma_j(x,\xi;\theta) \leq \frac{C_\varepsilon \exp\left(\frac{\xi_j}{1-\rho}\right)}{J}, \qquad \frac{\int \sigma_j(\nu)\sigma_0(\nu)dF(\nu)}{\sigma_j(x,\xi;\theta)} \geq \frac{1}{C_\varepsilon}. \tag{16}$$

Proof. Set $U = \frac{1}{J}\sum_{j\in\mathcal{J}_+} \exp\left(\frac{\xi_j}{1-\rho}\right)$ and choose B with $\Pr(\|\nu\|_2 > B) < \frac{\varepsilon}{2}$. On $\|\nu\|_2 \leq B$, compactness gives $|x_j'\beta + x_j'\Sigma\nu| \leq \widetilde{C}$, and hence $\widetilde{C}^{-1}U^{1-\rho} \leq D(\nu)^{1-\rho} \leq \widetilde{C}U^{1-\rho}$. Therefore,

$$\varepsilon \leq \sigma_0(x,\xi;\theta) \leq \frac{\varepsilon}{2} + \frac{1}{1+\widetilde{C}^{-1}U^{1-\rho}}, \qquad \varepsilon \leq 1-\sigma_0(x,\xi;\theta) \leq \frac{\varepsilon}{2} + \widetilde{C}U^{1-\rho}.$$

Thus $\widetilde{C}^{-1} \leq U \leq \widetilde{C}$. On any fixed bounded taste set, after increasing $\widetilde{C}$ if necessary,

$$\frac{\exp\left(\frac{\xi_j}{1-\rho}\right)}{\widetilde{C}J} \leq \sigma_j(\nu) \leq \frac{\widetilde{C}\exp\left(\frac{\xi_j}{1-\rho}\right)}{J}, \qquad \sigma_0(\nu) \geq \frac{1}{\widetilde{C}}.$$

Integrating over $\|\nu\|_2 \leq B$ gives (14).

Now let $x \in \mathcal{D}_J$ and $\min_j J\sigma_j(x,\xi;\theta) \geq \varepsilon$. Choose $h = \frac{1}{1+\|\Sigma\nu\|_2 \sup_{u,v\in\mathcal{X}}\|u-v\|_2}$. Cauchy–Schwarz and compactness of Θ give

$$\frac{1}{J}\sum_{k\in\mathcal{J}_+} \mathbf{1}(|(x_k-x_j)'\Sigma\nu| \leq 1) \geq \frac{c_x}{\left(1+\|\Sigma\nu\|_2 \sup_{u,v\in\mathcal{X}}\|u-v\|_2\right)^\alpha} \geq \frac{1}{\widetilde{C}(1+\|\nu\|_2)^\alpha}.$$

Choose j minimizing ξ_j. For every k satisfying $|(x_k-x_j)'\Sigma\nu| \leq 1$, $\delta_k(\nu)-\delta_j(\nu) \geq -\widetilde{C}$, since $\xi_k \geq \xi_j$ and the mean utilities are bounded. Then, we have $\sigma_j(\nu) \leq \sigma_{j|1}(\nu) \leq \frac{\widetilde{C}(1+\|\nu\|_2)^\alpha}{J}$. Choose a fixed B large enough that $\widetilde{C}\int_{\|\nu\|_2>B}(1+\|\nu\|_2)^\alpha dF(\nu) \leq \frac{\varepsilon}{2}$. Combining it with the bounded-taste upper bound gives

$$\frac{\varepsilon}{J} \leq \sigma_j(x,\xi;\theta) \leq \frac{\varepsilon}{2J} + \frac{\widetilde{C}\exp\left(\frac{\xi_j}{1-\rho}\right)}{J}.$$

Hence $\min_j \xi_j \geq -\widetilde{C}$.

For any inside j, the same product count now uses $\xi_k \geq -\widetilde{C}$ and gives

$$\sigma_j(\nu) \leq \frac{\widetilde{C}\exp\left(\frac{\xi_j}{1-\rho}\right)(1+\|\nu\|_2)^\alpha}{J}.$$

Multiply by $(1+\|\nu\|_2)^4$, integrate, and divide by the lower share bound in (14). This gives (15). Integrating without the fourth-power factor gives the upper share bound in (16).

Combining it with the integrated lower bound in (14) gives its last inequality. □

Lemma 11. *Under Assumption 4, consider case (RC1) with bounded ν or case (RC3). For both RC logit and RC nested logit, at $\xi = \xi^{\mathrm{BLP}}(x, \theta, s)$ with $s \in \mathcal{S}_J^+$,*

$$\|\xi^{\mathrm{BLP}}(x,\theta,s)\|_\infty \le \widetilde{C}, \qquad \inf_\nu \sigma_0(\nu) \ge \frac{1}{\widetilde{C}}, \qquad \sup_\nu \max_{j\in\mathcal{J}_+} \sigma_j(\nu) \le \frac{\widetilde{C}}{J}.$$

For finite mixtures the infimum and supremum are over m.

Proof. Compactness and bounded ν give $|x_j'\beta + x_j'\Sigma\nu| \le B < \infty$ uniformly; for finite mixtures, compactness gives $|x_j'\beta_m| \le B$. Evaluate the probabilities at $\xi = \xi^{\mathrm{BLP}}(x, \theta, s)$. For every consumer type, $\xi_j - \xi_k \ge 2B$ implies $\sigma_j(\nu) \ge \exp(\xi_j - \xi_k - 2B)\sigma_k(\nu)$. Integrating and using $\frac{s_j}{s_k} \le \frac{4C}{c}$ gives

$$\max_j \xi_j - \min_j \xi_j \le 2B + \log\left(\frac{4C}{c}\right).$$

Note that:

$$\exp(\min_j \xi_j - B) \le \frac{\sigma_{\mathrm{buy}}(\nu)}{\sigma_0(\nu)} \le \exp(\max_j \xi_j + B).$$

Since $s_0 \ge \frac{c_0}{2}$ and $1 - s_0 \ge \frac{c}{2}$, these inequalities imply

$$\min_j \xi_j \le B + \log\left(\frac{2}{c_0} - 1\right), \qquad \max_j \xi_j \ge \log\left(\frac{c}{2-c}\right) - B.$$

Then, we have $\|\xi\|_\infty \le \widetilde{C}$. Therefore, $\sigma_0(\nu) \ge \frac{1}{\widetilde{C}}$, and $\sigma_{j|1}(\nu) \le \frac{\widetilde{C}}{J}$. □

Proposition 3. *Under Assumption 4, RC logit and RC nested logit satisfy Assumptions 1–3 and 5(i)–5(iii) in cases (RC1)–(RC3), with $x \in \mathcal{D}_J$ in case (RC2). Under Assumption 6, Assumption 5(iv) also holds.*

Proof. RC logit is the case $\rho = 0$, or $\rho_m = 0$ for every type in a finite mixture. We therefore prove the result for RC nested logit.

Existence of the inverse. The taste distributions have finite moments of every order and $1 - \rho \ge 1 - \rho_{\max} > 0$, so differentiation of the share formulas gives Assumption 1. Equation (11) gives Assumption 2. For fixed J and $s \in \mathrm{int}(\Delta^J)$, consider

$$\int \log(1 + D(\nu)^{1-\rho}) dF(\nu) - \sum_{j\in\mathcal{J}_+} s_j \xi_j.$$

For finite mixtures, replace the integral by $-\sum_m \pi_m \log \sigma_0^{(m)}(x, \xi; \beta_m, \rho_m)$. Its gradient is the inside-share vector minus s_+, and its Hessian is positive definite by (11). Since $\log(1 +$

$D(\nu)^{1-\rho}) \geq \max(0, \max_j \delta_j(\nu))$, integrability of the utility indices gives the lower bound

$$\max(0, \max_j \xi_j) - \sum_{j \in \mathcal{J}_+} s_j \xi_j - C_J \geq \min_{j \in \mathcal{J}} s_j \, \|\xi\|_\infty - C_J, \qquad C_J < \infty.$$

The function is coercive and strictly convex, so its unique minimizer gives the inverse and hence Assumption 3.

Compact support and finite mixtures. Lemma 11 gives the inverse bound and the condition in Lemma 8.

Gaussian and lognormal tastes. Fix $\varepsilon \in (0, \frac{1}{2})$ and consider ξ with $\varepsilon \leq \sigma_0(x, \xi; \theta) \leq 1 - \varepsilon$. The proof of Lemma 10 bounds $\frac{1}{J} \sum_j \exp\left(\frac{\xi_j}{1-\rho}\right)$ above and away from zero. For Gaussian tastes, the within-nest formula therefore gives

$$\sigma_j(\nu) \leq \frac{\widetilde{C} \exp\left(\frac{\xi_j}{1-\rho}\right)}{J} \exp(\widetilde{C}(1 + \|\nu\|_2)).$$

Integration and (14) give the bounds in (16). For lognormal tastes, Lemma 10 gives these bounds when $x \in \mathcal{D}_J$ and $\min_j J\sigma_j \geq \varepsilon$, since all polynomial taste moments are finite. At $\xi = \xi^{\text{BLP}}(x, \theta, s)$, $s_0 \geq \frac{c_0}{2}$, $1 - s_0 \geq \frac{c}{2}$, and $Js_j \geq \frac{c}{2}$, so a fixed ε satisfies all these restrictions. Together with $s_j \leq \frac{2C}{J}$, (16) gives $\|\xi^{\text{BLP}}(x, \theta, s)\|_\infty \leq \widetilde{C}$ and the condition in Lemma 8.

The inverse bounds give Assumption 5(i). Lemmas 8 and 9 give Assumptions 5(ii)–5(iii) in all three cases.

Regularized separation. In case (RC1), $\log D(\nu)$ and $\log(1 + D(\nu)^{1-\rho})$ are convex jointly in (ξ, ν). Hence

$$\log \sigma_j(\nu) = \frac{\delta_j(\nu)}{1-\rho} - \rho \log D(\nu) - \log(1 + D(\nu)^{1-\rho}), \qquad \log \sigma_0(\nu) = -\log(1 + D(\nu)^{1-\rho})$$

are concave. Integration against the log-concave density preserves log-concavity by Prékopa (1973). Lemma 1 and Assumption 6 give Assumption 5(iv).

For cases (RC2)–(RC3), take any $\widetilde{\xi} \in \mathcal{A}_N(x, \theta, s)$ with $s \in \mathcal{S}_J^+$. Write $\sigma_j = \sigma_j(x, \widetilde{\xi}; \theta)$ and evaluate individual probabilities at $\widetilde{\xi}$. On a fixed bounded taste set, the utility indices at $\xi = 0$ are uniformly bounded. Integration gives $\sigma_0(x, 0; \theta) \geq \frac{1}{\widetilde{C}}$ and $\sigma_j(x, 0; \theta) \geq \frac{1}{\widetilde{C}J}$ for inside j. Comparing likelihood values gives

$$\sum_{j \in \mathcal{J}} s_j \log \frac{s_j}{\sigma_j} + \frac{\lambda_{N,J}}{2} \|\widetilde{\xi}\|_2^2 \leq \sum_{j \in \mathcal{J}} s_j \log \frac{s_j}{\sigma_j(x, 0; \theta)} \leq \widetilde{C}.$$

Grouping the inside choices gives

$$-s_0 \log \sigma_0 - (1 - s_0) \log(1 - \sigma_0) \leq \sum_{j \in \mathcal{J}} s_j \log \frac{s_j}{\sigma_j} + \log 2 \leq \widetilde{C}.$$

Since $s_0 \geq \frac{c_0}{2}$ and $1 - s_0 \geq \frac{c}{2}$, there is a fixed $\varepsilon > 0$ such that $\varepsilon \leq \sigma_0 \leq 1 - \varepsilon$.

Let $\sigma_+ = (\sigma_1, \ldots, \sigma_J)'$. The inside-share Jacobian is symmetric, with nonpositive off-diagonal entries and row sums $\int \sigma_j(\nu)\sigma_0(\nu)dF(\nu) > 0$. Adding up the share derivatives gives the FOC

$$\partial_\xi L_s(x, \widetilde{\xi}; \theta) = D_\xi \sigma_+(x, \widetilde{\xi}; \theta) \left(\frac{s_j}{\sigma_j} - \frac{s_0}{\sigma_0} \right)_{j \in \mathcal{J}_+} = \lambda_{N,J} \widetilde{\xi}. \tag{17}$$

In case (RC2), the outside-share bounds allow us to apply (14) at $\widetilde{\xi}$. Choose an inside product maximizing $\frac{s_j}{\sigma_j}$. If this ratio exceeds $\frac{s_0}{\sigma_0}$, the Jacobian signs and positive row sums in (17) give $\partial_{\xi_j} L_s > 0$, and hence $\widetilde{\xi}_j > 0$. Equation (14) then gives $\sigma_j \geq \frac{1}{\widetilde{C}J}$, so $\frac{s_j}{\sigma_j} \leq \widetilde{C}$. Otherwise every ratio is bounded by $\frac{s_0}{\sigma_0} \leq \widetilde{C}$. Since $s_j \geq \frac{c}{2J}$, all fitted inside shares satisfy $\sigma_j \geq \frac{1}{\widetilde{C}J}$.

Lemma 10 now gives (16) at $\widetilde{\xi}$, taking a smaller fixed ε if needed. Its upper share bound gives $\widetilde{\xi}_j \geq -\widetilde{C}$. Its last inequality also gives

$$\partial_{\xi_j} L_s(x, \widetilde{\xi}; \theta) \leq \frac{s_j}{1 - \rho} - \frac{s_0}{\sigma_0} \int \sigma_j(\nu)\sigma_0(\nu)dF(\nu) \leq \frac{\widetilde{C}}{J} - \frac{\sigma_j}{\widetilde{C}}. \tag{18}$$

If $\widetilde{\xi}_j \geq 0$, the FOC bounds σ_j above by $\frac{\widetilde{C}}{J}$. The lower share bound in (16) then bounds $\widetilde{\xi}_j$ above. Hence $\|\widetilde{\xi}\|_\infty \leq \widetilde{C}$.

In case (RC3), comparison with the BLP inverse gives $\|\widetilde{\xi}\|_2^2 \leq \|\xi^{\text{BLP}}(x, \theta, s)\|_2^2 \leq \widetilde{C}J$ by Lemma 11. Thus at least half the coordinates have bounded absolute value. The type-specific denominators are consequently bounded below, and every type's conditional inside probability for product j is at most $\frac{\widetilde{C}\exp(\widetilde{\xi}_j)}{J}$ whenever $\widetilde{\xi}_j \leq 0$. Equation (11) gives $\partial_{\xi_j} L_s \geq \frac{c}{2J} - \frac{\widetilde{C}\exp(\widetilde{\xi}_j)}{J}$, so the FOC bounds $\widetilde{\xi}_j$ below.

Since $\sum_m \pi_m \sigma_0^{(m)} = \sigma_0 \geq \frac{1}{\widetilde{C}}$, some type satisfies $\pi_m \sigma_0^{(m)} \geq \frac{1}{\widetilde{C}M}$. Its probability and outside share are bounded below. Its nested-logit formula then gives $\sum_m \pi_m \sigma_j^{(m)} \sigma_0^{(m)} \geq \frac{\exp(\widetilde{\xi}_j)}{\widetilde{C}J}$ when $\widetilde{\xi}_j \geq 0$. The first inequality in (18), with $\rho_{\max}$ in place of ρ, therefore bounds $\widetilde{\xi}_j$ above. Thus $\|\widetilde{\xi}\|_\infty \leq \widetilde{C}$. All type-specific outside shares are bounded below and inside shares are of order $\frac{1}{J}$.

In both cases, the inside-share Jacobian has row sums at least $\frac{1}{\widetilde{C}J}$. The row argument in

Lemma 8 bounds its inverse by $\widetilde{C}J$, so (17) gives

$$\max_{j\in\mathcal{J}_+}\left|\frac{s_j}{\sigma_j}-\frac{s_0}{\sigma_0}\right|\leq\widetilde{C}J\lambda_{N,J}\|\widetilde{\xi}\|_\infty.$$

Since $\sum_{j\in\mathcal{J}}\sigma_j\frac{s_j}{\sigma_j}=1$, the same bound, with a larger constant, holds for $\max_{j\in\mathcal{J}}|\frac{s_j}{\sigma_j}-1|$. With $C_\star$ chosen to cover this bound, Assumption 6 makes it at most $\frac{1}{2}$. Hence $\sigma_j\leq\frac{\widetilde{C}}{J}$ and $J\max_{j\in\mathcal{J}_+}|\sigma_j-s_j|\leq\widetilde{C}J\lambda_{N,J}\|\widetilde{\xi}\|_\infty$. The inverse calculations above also apply with any fixed wider share bounds. Integrating the inside-share inverse Jacobian along $s+u(\sigma-s)$, $0\leq u\leq 1$, therefore yields

$$\|\widetilde{\xi}-\xi^{\mathrm{BLP}}(x,\theta,s)\|_\infty\leq\widetilde{C}J\lambda_{N,J}\|\widetilde{\xi}\|_\infty\leq\widetilde{C}J\lambda_{N,J}. \tag{19}$$

With the common $C_\star$ chosen also to dominate these recovery bounds, Assumption 6 places every maximizer within distance $\frac{r_0}{2}$ of the BLP inverse, where Lemma 3 gives uniqueness. For fixed N,J, (x,s,θ) ranges over a compact set, using $\mathcal{D}_J$ in case (RC2), and $-L_s^{\mathrm{R}}$ is uniformly coercive. If the gap in Assumption 5(iv) had infimum zero, take a convergent parameter sequence and maximizers subject to $\|\xi-\xi^{\mathrm{BLP}}(x,\theta,s)\|_\infty\geq r_0$. Coercivity bounds these maximizers, and continuity gives a limiting global maximizer at distance at least r_0 from the limiting inverse. This contradicts the bound $\frac{r_0}{2}$. □

B.3 Parameter derivatives

Lemma 12. *Under Assumption 4, Assumption 9(ii) holds for simple logit and the RC logit and RC nested-logit models in cases (RC1)–(RC3), with $x\in\mathcal{D}_J$ in case (RC2).*

Proof. **Continuous tastes.** Use the conditional expectations $\mathbb{E}_j$ and individual scores h_{jk} from the proof of Lemma 9. For the coefficient parameters, compactness gives $\|\partial_\theta(x_j'\beta+x_j'\Sigma\nu)\|_2\leq\widetilde{C}(1+\|\nu\|_2)$, so $\|\partial_\theta\log\sigma_j(\nu)\|_2\leq\widetilde{C}(1+\|\nu\|_2)$. Its first ξ_k derivative and the row sum of its second ξ derivatives are bounded by $\widetilde{C}(1+\|\nu\|_2)(\mathbf{1}(j=k)+\sigma_{k|1}(\nu))$ for inside j, and by $\widetilde{C}(1+\|\nu\|_2)\sigma_k(\nu)$ for $j=0$.

Throughout the r_0-ball, (12) bounds the aggregate outside and purchase shares away from zero and gives $\min_j J\sigma_j\geq\frac{1}{\widetilde{C}}$. For lognormal tastes and $x\in\mathcal{D}_J$, Lemma 10 therefore gives (15); the outside bound follows from $\mathbb{E}(1+\|\nu\|_2)^4<\infty$. For bounded tastes the same bound is immediate. For Gaussian tastes, the inverse bounds and the probability comparison leading to (12) give $\sigma_k(\nu)\leq\frac{\widetilde{C}}{J}\exp(\widetilde{C}(1+\|\nu\|_2))$; integration and the aggregate share lower

bounds give the same moment bound. Thus, for $1 \le p \le 4$ and inside k,

$$\max_{j\in\mathcal{J}} \mathbb{E}_j(1+\|\nu\|_2)^4 \le \widetilde{C}, \qquad \int (1+\|\nu\|_2)^p \sigma_k(\nu) dF(\nu) = \sigma_k(x,\xi;\theta)\mathbb{E}_k(1+\|\nu\|_2)^p \le \frac{\widetilde{C}}{J}.$$

Equation (13) gives

$$\partial^2_{\xi_k\theta'} \log \sigma_j(x,\xi;\theta) = \mathbb{E}_j[\partial^2_{\xi_k\theta'} \log \sigma_j(\nu)] + \mathrm{Cov}_j(h_{jk}, \partial_{\theta'} \log \sigma_j(\nu)).$$

The individual bounds control this expression and its ξ directional derivative by a fixed multiple of $\mathbf{1}(j=k) + \mathbb{E}_j[(1+\|\nu\|_2)\sigma_{k|1}] + \mathbb{E}_j\sigma_{k|1}$. For $j=0$, replace $\sigma_{k|1}$ by σ_k and omit the indicator. Summing with weights s_j and using (12) gives

$$\max_k \sum_{j\in\mathcal{J}} s_j \left(\|\partial^2_{\xi_k\theta'} \log \sigma_j\|_2 + \|D_\xi \partial^2_{\xi_k\theta'} \log \sigma_j[w]\|_2 \right) \le \frac{\widetilde{C}}{J}, \qquad \|w\|_\infty \le 1.$$

For the nesting parameter, write $\bar{\delta}(\nu) = \sum_j \sigma_{j|1}(\nu)\delta_j(\nu)$ and $H(\nu) = -\sum_j \sigma_{j|1}(\nu) \log \sigma_{j|1}(\nu)$, with sums over $\mathcal{J}_+$. Holding β, Σ, ξ, ν fixed gives

$$\partial_\rho \log \sigma_j(\nu) = \frac{\delta_j(\nu) - \bar{\delta}(\nu)}{(1-\rho)^2} + \sigma_0(\nu)(\log J - H(\nu)), \qquad \partial_\rho \log \sigma_0(\nu) = \sigma_{\text{buy}}(\nu)(H(\nu) - \log J).$$

The inverse bounds imply $\max_{j,k} |\delta_j(\nu) - \delta_k(\nu)| \le \widetilde{C}(1+\|\nu\|_2)$, and

$$0 \le \log J - H(\nu) \le \frac{\max_j \delta_j(\nu) - \min_j \delta_j(\nu)}{1-\rho} \le \widetilde{C}(1+\|\nu\|_2).$$

Thus $|\partial_\rho \log \sigma_j(\nu)| \le \widetilde{C}(1+\|\nu\|_2)$. Using $\partial_{\xi_k} H(\nu) = \frac{\sigma_{k|1}(\nu)(\bar{\delta}(\nu) - \delta_k(\nu))}{(1-\rho)^2}$ and the derivatives in Lemma 9 gives

$$|\partial^2_{\xi_k\rho} \log \sigma_j(\nu)| + \sum_{\ell\in\mathcal{J}_+} |\partial^3_{\xi_\ell\xi_k\rho} \log \sigma_j(\nu)| \le \widetilde{C}(1+\|\nu\|_2)(\mathbf{1}(j=k) + \sigma_{k|1}(\nu)).$$

For $j=0$ the RHS is $\widetilde{C}(1+\|\nu\|_2)\sigma_k(\nu)$. Applying (13) and the preceding bounds gives the mixed bounds for ρ.

Finite mixtures. Lemma 11 and the probability comparisons leading to (12) give $\sigma_j^{(m)} \le \frac{\widetilde{C}}{J}$ for inside products and $\frac{\sigma_j^{(m)}}{\sigma_j} \le \widetilde{C}$ for every category and type. The preceding calculations therefore apply to each β_m and ρ_m, with bounded utility and nesting scores. For

π_m, $m < M$, we have

$$\partial_{\pi_m} \log \sigma_j = \frac{\sigma_j^{(m)} - \sigma_j^{(M)}}{\sigma_j}.$$

Differentiating gives

$$\partial^2_{\xi_k \pi_m} \log \sigma_j = \frac{\sigma_j^{(m)}}{\sigma_j} \left(\partial_{\xi_k} \log \sigma_j^{(m)} - \partial_{\xi_k} \log \sigma_j \right) - \frac{\sigma_j^{(M)}}{\sigma_j} \left(\partial_{\xi_k} \log \sigma_j^{(M)} - \partial_{\xi_k} \log \sigma_j \right).$$

The bounded ratios, individual derivative bounds, and their directional derivatives imply

$$|\partial^2_{\xi_k \pi_m} \log \sigma_j| + |D_\xi \partial^2_{\xi_k \pi_m} \log \sigma_j[w]| \leq \widetilde{C} \left(\mathbf{1}(j = k) + \frac{1}{J} \right), \qquad j \in \mathcal{J}_+.$$

For $j = 0$ the RHS is $\frac{\widetilde{C}}{J}$. Summing with weights s_j gives the J^{-1} bounds for C_s and $D_\xi C_s[w]$. Integration along the path between ξ and ξ' gives the Lipschitz bound. □