Convergence properties of BLP's fixed-point iteration

Miguel Zerecero*

University of California, Irvine

September 22, 2025

Abstract

This paper revisits the BLP inversion and clarifies a common misconception: the fixed-point iteration used to recover mean utilities is *not* a contraction mapping; Berry, Levinsohn, and Pakes (1995) proved it only for a componentwise bounded (truncated) operator. Nevertheless, I show that iterating the unbounded operator converges to the unique fixed point. Thus, practitioners can continue to use the standard unbounded iteration with formal convergence guarantees. I also provide bounds linking the local convergence rate to the outside share, explaining the empirical observation of faster convergence when the outside option is large.

JEL Codes: C25, L13

Keywords: Random coefficients logit demand, BLP, Contraction mapping, Convergence

In their seminal paper, Berry, Levinsohn, and Pakes (1995) (BLP henceforth) introduce the random-coefficient logit model and an estimator for it. It has been the workhorse for discrete-choice demand estimation for the past 30 years. A key step is an inversion that, for a given distribution of random coefficients θ, solves for the mean utilities δ via the fixed-point iteration

$$\delta \leftarrow \delta + \log \hat{\mathbf{s}} - \log \mathcal{S}(\delta; \theta), \tag{1}$$

*Contact: mzerecer@uci.edu. I thank Jiawei Chen and Max Lesellier for their comments and suggestions. Any remaining errors are my own.

where log acts elementwise on the observed share vector $\hat{\mathbf{s}}$, and $\mathcal{S}(\delta;\theta)$ is the model-implied share vector. Let $T(\delta)$ denote the right-hand side of (1), and its Jacobian as $\nabla T(\delta) := \partial T(\delta)/\partial\delta$.

BLP state in their main text (p. 865) that $T(\delta)$, as defined in (1), is a contraction mapping. This is incorrect. In fact, they prove a contraction mapping only for a *bounded* operator

$$\overline{T}(\delta) \equiv \min\{T(\delta), \overline{\delta}\}, \tag{2}$$

where the minimum is taken componentwise and $\overline{\delta}$ is an upper bound on δ. To the best of my knowledge, the literature has not provided theoretical guarantees that the *unbounded* iteration (1) converges to the fixed point, and has wrongly claim that BLP showed (1) is a contraction mapping.[1]

In this note I prove that the unbounded iteration (1)—the procedure virtually all practitioners use—converges to the unique fixed point, even though $T(\delta)$ is not a contraction mapping. Thus, the usual practice is theoretically justified.

The argument applies a result of Bifulco, Glück, Krebs, and Kukharskyy (2025) that upgrades *local* attractiveness of a fixed point to *global* attractiveness for nonexpansive maps. Since $T(\delta)$ is nonexpansive and its fixed point is locally attractive, convergence follows. Unlike BLP, no truncation or boundedness is required.

In addition, I discuss the empirical observation that convergence is faster when the outside good market share is larger. Dubé, Fox, and Su (2012, Table III) document that larger outside shares are associated with smaller Lipschitz constants. However, the Lipschitz constant is a global bound defined over the entire set $\mathbb{R}^J$. If we follow BLP, using the sup-norm $\|\cdot\|_\infty$, we have that $\|\nabla T(\delta)\|_\infty \leq 1$ everywhere, but one cannot obtain a global bound that is strictly less than 1. This is precisely why BLP imposed boundedness: on a compact set, a maximizer of $\|\nabla T(\delta)\|_\infty$ exists and can be strictly smaller than 1. Thus, the Lipschitz constants reported in Dubé et al. (2012) must have been computed for a restricted subset of $\mathbb{R}^J$, most likely at or near

[1] Without claiming to be an exhaustive list, here are some prominent examples. First, in their review of best practices for BLP estimation, Conlon and Gortmaker (2020, p. 1119) write that BLP showed (1) is a contraction mapping. Second, Train (2009, pp. 322–323) states that BLP proved that iterating (1) converges to the fixed point. Third, Nevo (2000, pp. 532–533) describes (1) as a "contraction mapping" and refers to BLP for a proof of convergence. Fourth, Dubé, Fox, and Su (2012) refer to (1) as "BLP's nested contraction-mapping algorithm," and similar. Finally, in deriving a "damped" operator for the nested logit version of BLP, the appendix of Grigolon and Verboven (2014) discusses conditions for a contraction mapping à la BLP but omits the truncation that BLP use to obtain a Lipschitz constant strictly less than one.

the fixed point δ^*.

Nevertheless, the Jacobian evaluated at the fixed point $\nabla T(\delta^*)$, remains informative about convergence. The spectral radius of $\nabla T(\delta^*)$ governs the local convergence rate, and is bounded above by $\|\nabla T(\delta^*)\|_\infty$. This allows us to relate convergence speed to the outside good market share: larger outside shares imply smaller values of $\|\nabla T(\delta^*)\|_\infty$, and hence faster local convergence. I show that without distributional assumptions on the random coefficients this bound involves the outside share plus a covariance term. Under the common assumption of normally distributed random coefficients, the bound simplifies and depends primarily on the observed outside good share.

1 The random coefficient logit model

There are $J+1$ products, including the outside option. Each consumer i derives utility from product j given by

$$u_{ij} = \delta_j + \mu_{ij} + \varepsilon_{ij},$$

where δ_j is the mean utility of product j, μ_{ij} captures deviations due to heterogeneous tastes, and ε_{ij} is an i.i.d. Type I extreme value error. The outside option is normalized so that $u_{i0} = \varepsilon_{i0}$.

Consumers choose the product that yields the highest utility. Aggregating over the distribution of heterogeneous tastes yields the market share function

$$\mathcal{S}_j(\delta;\theta) = \int \frac{\exp(\delta_j + \mu_{ij})}{1 + \sum_{k=1}^{J} \exp(\delta_k + \mu_{ik})} dF(\boldsymbol{\mu}_i;\theta),$$

where $F(\boldsymbol{\mu}_i;\theta)$ denotes the distribution of consumer heterogeneity, parameterized by θ. In practice, this integral is evaluated by simulation.

The empirical problem is to recover the mean utilities δ that rationalize the observed market shares $\hat{\mathbf{s}}$. This inversion step is central to the BLP estimator, since δ can then be related linearly to observed product characteristics, prices, and unobserved demand shocks.

What did BLP show? In their Appendix, BLP establish, first, that the bounded operator $\overline{T}(\delta)$, as defined by (2), is a contraction mapping. By Banach's fixed point theorem, this implies (i) existence and uniqueness of the solution to the bounded system $\delta = \overline{T}(\delta)$, and (ii) that iterates of $\overline{T}(\delta)$ converge to the fixed point. BLP also show that the solution to the bounded system cannot lie on the boundary.[2] Thus, the solution to the bounded system is the same as the one for the unbounded system $\delta = T(\delta)$. This implies existence and uniqueness of the solution for the unbounded system. But this *does not* imply that iterates over the unbounded operator $T(\delta)$ converge to the fixed point. We need a different argument for that.[3]

2 Convergence properties of the unbounded operator

I present the main results here and explain briefly the proofs. I leave all the detailed proofs for the Appendix.

Proposition 1. *The map $T(\delta)$ in (1) is not a contraction mapping on $\mathbb{R}^J$. Nevertheless, the iteration $\delta^{t+1} = T(\delta^t)$ converges to the unique fixed point δ^* from any starting value.*

The proof for Proposition 1 is an application of a Theorem by Bifulco et al. (2025) that extends the local attractiveness of the fixed point to be global for non-expansive Lipschitz mappings. Thus, it suffices to prove that around the fixed point, $T(\delta)$ is locally attractive and Lipschitz continuous with Lipschitz constant ≤ 1.

The asymptotic local (linear) rate of convergence is bounded above by the spectral radius of the Jacobian at the fixed point, denoted $\rho\big(\nabla T(\delta^*)\big)$, which is in turn bounded above by $\|\nabla T(\delta^*)\|_\infty$.[4] The following propositions relate the local rate of convergence to the observed outside market share.

Before presenting the following propositions, let me introduce some auxiliary notation. Let

$$\pi_{ij} := \frac{\exp(\delta_j + \mu_{ij})}{1 + \sum_{k=1}^{J} \exp(\delta_k + \mu_{ik})}, \quad \text{and} \quad f_{ij} := \frac{\pi_{ij}}{\mathcal{S}_j(\delta;\theta)}.$$

---

[2] The proof in their published paper has a couple of typographic errors related to the definition of bounds. These errors are not present in the slightly different NBER working paper proof.

[3] Let (X, d) be a metric space. Even if $d(T(x), T(y)) < d(x, y)$ for all distinct $x, y \in X$, T is not a contraction mapping because there need not exist a *uniform* $c < 1$ (the Lipschitz constant) with $d(T(x), T(y)) \leq c\, d(x, y)$ for all x, y. Without such a uniform constant one can have factors $c_t \in (0, 1)$ with $c_t \to 1$ so that $\prod_{k=0}^{t-1} c_k$ stays bounded away from zero and the distance does not vanish.

[4] See Ortega and Rheinboldt (1970, Ch. 10, Thm. 10.1.4).

The following proposition bounds the spectral radius without distribution assumptions on the distribution F.

Proposition 2. $\rho(\nabla T(\delta^*)) \leq (1 - \hat{s}_0) - \min_j \{\mathrm{Cov}(f_{ij}, \pi_{i0})\}$.

The bound links the local (linear) convergence rate, governed by $\rho(\nabla T(\delta^*))$, to the outside share $\hat{s}_0$. Larger $\hat{s}_0$ tightens the bound. The covariance term captures how the product–j weights f_{ij} comove with the outside choice probability π_{i0}. Without additional restrictions on F, it is hard to determine the sign of the covariance theoretically, so a sharper share-only bound is generally unavailable.

We can get a more explicit bound using the following assumption:

Assumption 1. *The random coefficient components μ_{ij} are independent across goods, and each are distributed normal with mean zero and variance σ_j^2.*

Proposition 3. *Let $\overline{\sigma}^2 = \max_j \sigma_j^2$. If Assumption 1 holds, then*

$$\rho(\nabla T(\delta^*)) \leq (1 - \hat{s}_0) + \sqrt{\left(\exp(\overline{\sigma}^2) - \hat{s}_0\right)(1 - \hat{s}_0)}.$$

We can show that $\rho(\nabla T(\delta^*)) \leq 1$ regardless of the distributional assumptions. So the bound above is only informative when $\hat{s}_0 > 1/2$, as $\exp(\overline{\sigma}^2) \geq 1$.

Random coefficient nested logit. By the same reasoning, the unbounded iteration for the nested-logit variant of BLP in Grigolon and Verboven (2014) is globally convergent.

3 Conclusion

This paper establishes global convergence of the unbounded BLP fixed-point iteration to the unique fixed point, closing the gap between BLP's bounded contraction argument and the algorithm used in practice. The result, together with the spectral-radius bounds that highlight the role of the outside share, provides the correct theoretical baseline for comparing this inversion method with competing alternatives, like the one of Li (2018).

APPENDIX

I begin with definitions and a few auxiliary lemmas (some standard; proofs included for completeness). I then state a theorem of Bifulco et al. (2025) and reproduce its proof for convenience. Proposition 1 follows by verifying that the unbounded operator $T(\delta)$ satisfies the theorem's conditions. The remaining propositions reduce to characterizing $\|\nabla T(\delta^*)\|_\infty$.

Proof of Proposition 1

Definition 1 (Lipschitz and non-expansive mappings)**.** *Let (X, d) be a metric space. A mapping $F : X \to X$ is* Lipschitz continuous *with* Lipschitz constant $L \geq 0$ *if*

$$d\big(F(x), F(y)\big) \;\leq\; L\, d(x, y) \quad \text{for all } x, y \in X.$$

If $L \leq 1$, we say F is non-expansive *(with respect to d).*

Definition 2 (Local asymptotic stability (Lyapunov))**.** *Let (X, d) be a metric space and let $x^* \in X$ be a fixed point of $F : X \to X$. We say x^* is* locally asymptotically stable *if there exists a neighborhood $\mathcal{U}$ of x^* such that for any $x^0 \in \mathcal{U}$, the iterates $x^{t+1} = F(x^t)$ satisfy $x^t \to x^*$ as $t \to \infty$.*

Lemma 1 (Spectral radius condition for local stability)**.** *Let $F : \mathbb{R}^J \to \mathbb{R}^J$ be continuously differentiable in a neighborhood of a fixed point x^*, and suppose the spectral radius satisfies $\rho(\nabla F(x^*)) < 1$. Then x^* is locally asymptotically stable for the iteration $x^{t+1} = F(x^t)$.*

Proof. Fix $\varepsilon > 0$ so small that $\rho(\nabla F(x^*)) + \varepsilon =: q < 1$. By a standard result from matrix analysis (see Lemma 5.6.10 of Horn and Johnson (2013)), there exists a vector norm $\|\cdot\|_*$ on $\mathbb{R}^J$ with induced matrix norm $\|\cdot\|_*$ such that

$$\|\nabla F(x^*)\|_* \;\leq\; \rho(\nabla F(x^*)) + \varepsilon \;=\; q \;<\; 1.$$

By continuity of ∇F, there is $r > 0$ and $q' < 1$ with

$$\sup_{z \in B_*(x^*, r)} \|\nabla F(z)\|_* \;\leq\; q' \;<\; 1,$$

where $B_*(x^*, r) := \{z : \|z - x^*\|_* \le r\}$. For any $x, y \in B_*(x^*, r)$, the mean-value (integral) formula gives

$$F(x) - F(y) \;=\; \int_0^1 \nabla F\big(y + t(x-y)\big)\,(x-y)\,dt,$$

whence

$$\|F(x) - F(y)\|_* \;\le\; \int_0^1 \|\nabla F(y + t(x-y))\|_*\,dt\;\|x - y\|_* \;\le\; q'\,\|x - y\|_*.$$

Thus F is a contraction on $B_*(x^*, r)$ (in the metric $d_*(x, y) = \|x - y\|_*$). Since $F(x^*) = x^*$, for any $x \in B_*(x^*, r)$,

$$\|F(x) - x^*\|_* \le q'\|x - x^*\|_* \le q'r < r,$$

so F maps $B_*(x^*, r)$ into itself. By Banach's fixed-point theorem, the iterates $x^{t+1} = F(x^t)$ converge to x^* for every $x^0 \in B_*(x^*, r)$. □

Remark 1 (Linearized stability)**.** *Lemma 1 is the discrete-time principle of linearized stability: if the spectral radius of the Jacobian at a fixed point is < 1, then the fixed point is locally asymptotically stable. See, e.g., Ortega and Rheinboldt (1970, Ch. 10).*

Lemma 2 (Non-expansiveness of the unbounded operator)**.** *Let $T(\delta) = \delta + \log \hat{\mathbf{s}} - \log \mathcal{S}(\delta; \theta)$ and equip $\mathbb{R}^J$ with the metric $d_\infty(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_\infty$. Then T is globally non-expansive:*

$$\|T(\delta) - T(\delta')\|_\infty \;\le\; \|\delta - \delta'\|_\infty \qquad \textit{for all } \delta, \delta'.$$

Moreover, with $\nabla T(\delta) := \partial T(\delta)/\partial\delta$, we have $\|\nabla T(\delta)\|_\infty \le 1$ for all δ, and the bound is tight.

Proof. By the mean-value theorem it suffices to bound $\|\nabla T(\delta)\|_\infty$. Let $\pi_{ij}(\delta, \boldsymbol{\mu}_i)$ denote individual logit choice probabilities (conditional on $\boldsymbol{\mu}_i$). Then

$$\mathcal{S}_j(\delta; \theta) \;=\; \int \pi_{ij}(\delta, \boldsymbol{\mu}_i)\, dF(\boldsymbol{\mu}_i; \theta).$$

Differentiating under the integral gives

$$\frac{\partial \log \mathcal{S}_j}{\partial \delta_m} \;=\; \frac{\displaystyle\int \pi_{ij}(\mathbf{1}\{j = m\} - \pi_{im})\, dF(\boldsymbol{\mu}_i; \theta)}{\displaystyle\int \pi_{ij}\, dF(\boldsymbol{\mu}_i; \theta)}.$$

Hence

$$\nabla T_{jm}(\boldsymbol{\delta}) \;=\; \delta_{jm} - \frac{\partial \log \mathcal{S}_j}{\partial \delta_m} \;=\; \frac{\displaystyle\int \pi_{ij}\,\pi_{im}\,dF(\boldsymbol{\mu}_i;\theta)}{\displaystyle\int \pi_{ij}\,dF(\boldsymbol{\mu}_i;\theta)} \;\geq\; 0.$$

For any fixed j,

$$\sum_{m=1}^{J} \nabla T_{jm}(\boldsymbol{\delta}) \;=\; \frac{\displaystyle\int \pi_{ij}\Big(\sum_{m=1}^{J}\pi_{im}\Big)\,dF(\boldsymbol{\mu}_i;\theta)}{\displaystyle\int \pi_{ij}\,dF(\boldsymbol{\mu}_i;\theta)} \;=\; \frac{\displaystyle\int \pi_{ij}\,(1-\pi_{i0})\,dF(\boldsymbol{\mu}_i;\theta)}{\displaystyle\int \pi_{ij}\,dF(\boldsymbol{\mu}_i;\theta)} \;\leq\; 1,$$

since $0 \leq \pi_{i0} \leq 1$. Therefore $\|\nabla T(\boldsymbol{\delta})\|_\infty = \max_j \sum_m \nabla T_{jm}(\boldsymbol{\delta}) \leq 1$. Tightness follows along sequences with $\pi_{i0} \to 0$. □

Remark 2. *Because the global non-expansive bound in Lemma 2 is tight (i.e., it cannot be improved to a uniform constant < 1 on $\mathbb{R}^J$), Berry et al. (1995) truncate the map and work on a compact set $K = [\underline{\delta}, \overline{\delta}] \subset \mathbb{R}^J$. On K one has*

$$\sup_{\delta \in K} \|\nabla T(\delta)\|_\infty = c < 1,$$

so the truncated operator $\overline{T} : K \to K$ is a contraction mapping and Banach's theorem yields existence, uniqueness, and convergence of the truncated iteration.

Lemma 3 (Spectral radius bounded by a matrix norm)**.** *Let $\|\cdot\|$ be any operator (induced) matrix norm on $\mathbb{R}^{J\times J}$. For any matrix $\mathbf{A} \in \mathbb{R}^{J\times J}$,*

$$\rho(\mathbf{A}) \;\leq\; \|\mathbf{A}\|.$$

Proof. Let λ be any eigenvalue of $\mathbf{A}$ with (nonzero) eigenvector $\mathbf{v}$. By definition of the induced norm,

$$\|\mathbf{A}\| \;=\; \sup_{x\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \;\geq\; \frac{\|\mathbf{A}\mathbf{v}\|}{\|\mathbf{v}\|} \;=\; \frac{\|\lambda\mathbf{v}\|}{\|\mathbf{v}\|} \;=\; |\lambda|.$$

Taking the supremum over all eigenvalues yields $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$. □

Lemma 4 (Local attractiveness at the fixed point)**.** *Let δ^* satisfy $\mathcal{S}(\delta^*;\theta) = \hat{\mathbf{s}}$ with $\hat{s}_0 \in (0,1)$ and $\hat{s}_j > 0$ for all $j \geq 1$. Then $\rho\big(\nabla T(\delta^*)\big) < 1$.*

Proof. By Lemma 2, for each j,

$$\sum_{m=1}^{J} \nabla T_{jm}(\delta^*) \;=\; \frac{\displaystyle\int \pi_{ij}(1-\pi_{i0})\, dF(\boldsymbol{\mu}_i;\theta)}{\displaystyle\int \pi_{ij}\, dF(\boldsymbol{\mu}_i;\theta)} \;=\; 1 - \frac{\displaystyle\int \pi_{ij}\,\pi_{i0}\, dF(\boldsymbol{\mu}_i;\theta)}{\hat{s}_j} \;<\; 1,$$

because $\pi_{ij} > 0$ and $\pi_{i0} > 0$ a.s. imply $\int \pi_{ij}\pi_{i0}\, dF > 0$. Thus $\|\nabla T(\delta^*)\|_\infty = \max_j \sum_m \nabla T_{jm}(\delta^*) < 1$ and hence, by Lemma 3, $\rho\big(\nabla T(\delta^*)\big) \leq \|\nabla T(\delta^*)\|_\infty < 1$. □

Theorem 1 (Theorem C.1 from Bifulco et al. (2025))**.** *Let (Z, d) be a metric space which is connected, i.e., it cannot be written as the union of two non-empty open disjoint subsets. Let $G : Z \to Z$ be a Lipschitz continuous function with Lipschitz constant 1, and let $z^* \in Z$ be a fixed point of G which is locally asymptotically stable.*

Then z^* is globally attractive, i.e., for each $z \in Z$ the sequence $G^n(z)$ converges to z^* as $n \to \infty$. In particular, z^* is the only fixed point of G.

Proof. Let B denote the basin of attraction of the fixed point z^*, i.e., the set of all $z \in Z$ for which $G^n(z) \to z^*$ as $n \to \infty$. Note that B is non-empty since $z^* \in B$.

Step 1: B is open. Since z^* is locally asymptotically stable, there exists $\delta > 0$ such that the ball $B_\delta(z^*) \subseteq B$. More generally, if $z_0 \in B$, then $G^n(z_0) \to z^*$, so eventually $G^n(z_0)$ gets arbitrarily close to z^*. By local asymptotic stability around z^*, points near $G^n(z_0)$ also converge to z^*. By continuity of G, points near z_0 will have their iterates near $G^n(z_0)$, so they also converge. Therefore B is open.

Step 2: B is closed. Let $(z_k)_{k\in\mathbb{N}}$ be a sequence in B that converges to some $z \in Z$. We need to show $z \in B$, i.e., $G^n(z) \to z^*$ as $n \to \infty$.

Let $\varepsilon > 0$ be arbitrary. Since $z_k \to z$, choose k_0 such that $d(z_{k_0}, z) < \varepsilon/2$.

Since $z_{k_0} \in B$, we have $G^n(z_{k_0}) \to z^*$ as $n \to \infty$. Therefore, there exists n_0 such that for all $n \geq n_0$:

$$d(G^n(z_{k_0}), z^*) < \varepsilon/2$$

Now, using the Lipschitz continuity with constant 1, for $n \geq n_0$:

$$\begin{aligned} d(G^n(z), z^*) &\leq d(G^n(z), G^n(z_{k_0})) + d(G^n(z_{k_0}), z^*) \\ &\leq d(z, z_{k_0}) + \varepsilon/2 \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon \end{aligned}$$

Since this holds for arbitrary $\varepsilon > 0$, we have $G^n(z) \to z^*$ as $n \to \infty$, so $z \in B$. Hence B is closed.

Step 3: Connected space. Since Z is connected and B is both open and closed (and non-empty), we must have $B = Z$. Therefore, z^* is globally attractive.

Step 4: Uniqueness. If there were two fixed points z_1^* and z_2^*, then since $B = Z$, we would have $G^n(z_1^*) \to z_2^*$. But $G^n(z_1^*) = z_1^*$ for all n, giving $z_1^* = z_2^*$. Therefore the fixed point is unique. □

Proof of Proposition 1. By the argument given on the main text, BLP showed that there exists a unique fixed point δ^*. By Lemma 2, T is non-expansive on $(\mathbb{R}^J, d_\infty)$, and as explained in Remark 2, $T(\delta)$ is not a contraction mapping on $\mathbb{R}^J$. By Lemma 4 and Lemma 1, the unique fixed point δ^* is locally asymptotically stable. Theorem 1 then gives global convergence of $\delta^{t+1} = T(\delta^t)$ to δ^* from any starting value. □

Proofs of Propositions 2 and 3

Lemma 5 (Variance of f_{ij})**.** *Let $f_{ij} := \pi_{ij}/\mathcal{S}_j(\delta;\theta)$. Then,*

(i) $\mathrm{Var}(f_{ij})$ *is nonincreasing in δ_j.*

(ii) *Under Assumption 1, for all δ_j (in particular at the fixed point δ^*),*

$$\mathrm{Var}(f_{ij}) \leq \frac{e^{\sigma_j^2}}{s_0} - 1.$$

Proof. Let $S_j := \mathcal{S}_j(\delta;\theta) = \mathbb{E}[\pi_{ij}]$ and recall $f_{ij} = \pi_{ij}/S_j$.

(i) $\mathrm{Var}(f_{ij})$ *is nonincreasing in* δ_j. Since $\mathbb{E}[f_{ij}] = 1$, we have

$$\mathrm{Var}(f_{ij}) = \mathbb{E}[f_{ij}^2] - 1, \qquad \frac{d}{d\delta_j}\mathrm{Var}(f_{ij}) = \frac{d}{d\delta_j}\mathbb{E}[f_{ij}^2] = 2\,\mathbb{E}\left[f_{ij}\,\frac{\partial f_{ij}}{\partial \delta_j}\right].$$

Using $\partial\pi_{ij}/\partial\delta_j = \pi_{ij}(1-\pi_{ij})$ and $\partial S_j/\partial\delta_j = \mathbb{E}[\pi_{ij}(1-\pi_{ij})]$,

$$\frac{\partial f_{ij}}{\partial \delta_j} = \frac{S_j\,\pi_{ij}(1-\pi_{ij}) - \pi_{ij}\,\mathbb{E}[\pi_{ij}(1-\pi_{ij})]}{S_j^2} = \frac{\pi_{ij}}{S_j^2}\Big(\mathbb{E}[\pi_{ij}^2] - S_j\,\pi_{ij}\Big).$$

Therefore

$$\mathbb{E}\left[f_{ij}\,\frac{\partial f_{ij}}{\partial \delta_j}\right] = \frac{1}{S_j^3}\Big(\mathbb{E}[\pi_{ij}^2]^2 - S_j\,\mathbb{E}[\pi_{ij}^3]\Big).$$

Recall $\pi_{ij} \in [0,1]$. Then, by Cauchy–Schwarz inequality,

$$\mathbb{E}[\pi_{ij}^2] = \mathbb{E}\big[\pi_{ij}^{3/2}\,\pi_{ij}^{1/2}\big] \;\le\; \sqrt{\mathbb{E}[\pi_{ij}^3]}\,\sqrt{\mathbb{E}[\pi_{ij}]}\,.$$

Then $\mathbb{E}[\pi_{ij}^2]^2 \le \mathbb{E}[\pi_{ij}^3]\mathbb{E}[\pi_{ij}]$, so $\dfrac{d}{d\delta_j}\mathrm{Var}(f_{ij}) \le 0$. Equality holds only if π_{ij} is constant, so the inequality is strict in any nondegenerate case.

(ii) Bound under Assumption 1. Write the "leave-j-out" sum $S_{-j,i} := \sum_{m\neq j} e^{\delta_m+\mu_{im}}$ and $B_{ij} := \frac{1}{1+S_{-j,i}} \in (0,1]$.

We have:

$$f_{ij} \;=\; \frac{\dfrac{e^{\delta_j}e^{\mu_{ij}}}{1+S_{-j,i}+e^{\delta_j+\mu_{ij}}}}{\mathbb{E}\left[\dfrac{e^{\delta_j}e^{\mu_{ij}}}{1+S_{-j,i}+e^{\delta_j+\mu_{ij}}}\right]} = \frac{\dfrac{e^{\mu_{ij}}}{1+S_{-j,i}+e^{\delta_j+\mu_{ij}}}}{\mathbb{E}\left[\dfrac{e^{\mu_{ij}}}{1+S_{-j,i}+e^{\delta_j+\mu_{ij}}}\right]}.$$

Then

$$\lim_{\delta_j\to-\infty} f_{ij} \;=\; \frac{\dfrac{e^{\mu_{ij}}}{1+S_{-j,i}}}{\mathbb{E}\left[\dfrac{e^{\mu_{ij}}}{1+S_{-j,i}}\right]}.$$

Moreover, under Assumption 1 (independence across goods),

$$\lim_{\delta_j \to -\infty} f_{ij} \;=\; \underbrace{\frac{e^{\mu_{ij}}}{\mathbb{E}[e^{\mu_{ij}}]}}_{=:X_{ij}} \cdot \underbrace{\frac{B_{ij}}{\mathbb{E}[B_{ij}]}}_{=:Y_{ij}}, \qquad X_{ij} \perp Y_{ij}.$$

Clearly $\mathbb{E}[X_{ij}] = \mathbb{E}[Y_{ij}] = 1$. Under Assumption 1, X_{ij} depends only on μ_{ij} while Y_{ij} depends only on $\{\mu_{im}\}_{m \neq j}$, so $X_{ij} \perp Y_{ij}$. Therefore

$$\lim_{\delta_j \to -\infty} \operatorname{Var}(f_{ij}) = \operatorname{Var}(X_{ij}) + \operatorname{Var}(Y_{ij}) + \operatorname{Var}(X_{ij})\ \operatorname{Var}(Y_{ij}).$$

If $\mu_{ij} \sim \mathcal{N}(0, \sigma_j^2)$, then

$$\operatorname{Var}(X_{ij}) = \frac{\mathbb{E}[e^{2\mu_{ij}}]}{\mathbb{E}[e^{\mu_{ij}}]^2} - 1 = \frac{e^{2\sigma_j^2}}{e^{\sigma_j^2}} - 1 = e^{\sigma_j^2} - 1.$$

Since $0 \leq B_{ij} \leq 1$, $\operatorname{Var}(B_{ij}) \leq \mathbb{E}[B_{ij}]\big(1 - \mathbb{E}[B_{ij}]\big)$, hence $\operatorname{Var}(Y_{ij}) \leq 1/\mathbb{E}[B_{ij}] - 1$. Combining,

$$\lim_{\delta_j \to -\infty} \operatorname{Var}(f_{ij}) \;\leq\; \Big(e^{\sigma_j^2} - 1\Big) + \Big(\tfrac{1}{\mathbb{E}[B_{ij}]} - 1\Big) + \Big(e^{\sigma_j^2} - 1\Big)\Big(\tfrac{1}{\mathbb{E}[B_{ij}]} - 1\Big) \;=\; \frac{e^{\sigma_j^2}}{\mathbb{E}[B_{ij}]} \;-\; 1.$$

Finally, since pointwise $B_{ij} = \pi_{i0}/(1 - \pi_{ij}) \geq \pi_{i0}$, taking expectations gives $\mathbb{E}[B_{ij}] \geq \mathbb{E}[\pi_{i0}] = s_0$. Using part (i), for any δ_j (in particular at δ^*),

$$\operatorname{Var}(f_{ij}) \;\leq\; \lim_{\delta_j \to -\infty} \operatorname{Var}(f_{ij}) \;\leq\; \frac{e^{\sigma_j^2}}{\mathbb{E}[B_{ij}]} \;-\; 1 \;\leq\; \frac{e^{\sigma_j^2}}{s_0} \;-\; 1.$$

$\square$

Proof of Proposition 2. From Lemma 2, for each j,

$$\sum_{m=1}^{J} \nabla T_{jm}(\delta^*) = \frac{\int \pi_{ij}(1 - \pi_{i0})\, dF}{\hat{s}_j} = 1 - \frac{\int \pi_{ij}\pi_{i0}\, dF}{\hat{s}_j}.$$

Note that
$$\frac{\int \pi_{ij}\pi_{i0}\,dF}{\hat{s}_j} = \mathbb{E}\left[\frac{\pi_{ij}}{\hat{s}_j}\pi_{i0}\right] = \hat{s}_0 + \operatorname{Cov}(f_{ij}, \pi_{i0}).$$

Thus for each j,
$$\sum_m \nabla T_{jm}(\delta^*) = (1 - \hat{s}_0) - \operatorname{Cov}(f_{ij}, \pi_{i0}).$$

Taking the maximum row sum,

$$\|\nabla T(\delta^*)\|_\infty = \max_j \sum_m \nabla T_{jm}(\delta^*) \leq (1 - \hat{s}_0) - \min_j \operatorname{Cov}(f_{ij}, \pi_{i0}).$$

Finally, Lemma 3 gives

$$\rho\big(\nabla T(\delta^*)\big) \leq \|\nabla T(\delta^*)\|_\infty \leq (1 - \hat{s}_0) - \min_j \operatorname{Cov}(f_{ij}, \pi_{i0}).$$

which is the desired bound. □

Proof of Proposition 3. Starting from Proposition 2 we get

$$\rho\big(\nabla T(\delta^*)\big) \leq (1 - \hat{s}_0) + \max_j |\operatorname{Cov}(f_{ij}, \pi_{i0})| \leq (1 - \hat{s}_0) + \max_j \sqrt{\operatorname{Var}(f_{ij})\ \operatorname{Var}(\pi_{i0})}$$

Since $0 \leq \pi_{i0} \leq 1$ with mean $\hat{s}_0$, its variance is bounded by

$$\operatorname{Var}(\pi_{i0}) \leq \hat{s}_0(1 - \hat{s}_0).$$

It remains to bound $\operatorname{Var}(f_{ij})$. Using Lemma 5, $\operatorname{Var}(f_{ij})$ is maximized when $\delta_j \longrightarrow -\infty$. Under Assumption 1, $\mu_{ij} \sim N(0, \sigma_j^2)$ independently across j. Again, using Lemma 5 we get, on the fixed point δ^*,
$$\operatorname{Var}(f_{ij}) \ \leq \ \frac{\exp(\sigma_j^2) - \hat{s}_0}{\hat{s}_0}.$$

Therefore
$$\rho\big(\nabla T(\delta^*)\big) \leq (1 - \hat{s}_0) + \max_j \sqrt{(\exp(\sigma_j^2) - \hat{s}_0)(1 - \hat{s}_0)}.$$

Letting $\overline{\sigma}^2 = \max_j \sigma_j^2$ gives the stated bound. □

References

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Bifulco, P., J. Glück, O. Krebs, and B. Kukharskyy (2025): "Single and Attractive: Uniqueness and Stability of Economic Equilibria Under Monotonicity Assumptions," Tech. rep., CESifo.

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

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

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

Horn, R. A. and C. R. Johnson (2013): *Matrix Analysis*, New York: Cambridge University Press, 2nd ed.

Li, L. (2018): "A general method for demand inversion," *arXiv preprint arXiv:1802.04444*.

Nevo, A. (2000): "A practitioner's guide to estimation of random-coefficients logit models of demand," *Journal of economics & management strategy*, 9, 513–548.

Ortega, J. and W. Rheinboldt (1970): *Iterative Solution of Nonlinear Equations in Several Variables*, vol. 30, SIAM.

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