Normalization, Convergence, and Speed in Demand and Gravity Inversion

Miguel Zerecero*

University of California, Irvine

August 28, 2026

Abstract

Many demand-estimation and gravity-type models in industrial organization, spatial, trade, and urban economics recover fundamentals by inverting observed allocations. These inversions share a common fixed-point structure and are typically solved by iteration. Because fundamentals are identified only up to a constant, the iteration imposes a seemingly innocuous normalization, which differs across fields: demand estimation fixes one alternative, the outside good, and drops its equation; gravity-type applications often update all equations symmetrically and normalize afterward. I show that both algorithms converge globally but can differ sharply in speed. The symmetric update is theoretically faster and avoids the well-known slow-convergence problem in demand estimation when the outside-good share is small. Simulations and standard benchmarks show substantial speed gains. For example, in the Nevo (2000) cereal application, the symmetric update reduces market-share function evaluations by more than 50%. It requires only minor changes to existing code.

JEL Codes: C25, C63, F10, L00, R12

Keywords: Demand inversion, BLP, gravity, fixed-point iteration, normalization, convergence

*First version: August 2026. Contact: mzerecer@uci.edu. This paper subsumes an earlier paper, first circulated in September 2025 under the title *Convergence Properties of BLP's Fixed-Point Iteration*. I thank David Agrawal, Jiawei Chen, Max Lesellier, and Jack Liebersohn for their comments and suggestions. Any remaining errors are my own.

1 Introduction

A common step in quantitative models is to recover fundamentals that rationalize an observed allocation as an equilibrium of the model. In the discrete-choice demand estimation literature (Berry, Levinsohn, and Pakes, 1995, BLP henceforth), this inversion step recovers mean product utilities from observed market shares. In quantitative trade, spatial, and urban models, it recovers location fundamentals from observed outcomes like output or employment. In both cases the inversion often has a logit or gravity structure.

The fundamentals are identified only up to a common constant, so their level must be normalized. In demand estimation, researchers usually fix one alternative, the outside good, by setting its utility to zero. The standard BLP iteration then drops the outside-good equation and updates only the remaining product utilities. I call this the *reference-fixed* iteration. Trade, spatial, and urban applications often have no obvious reference location. Researchers instead use aggregate normalizations, like dividing location fundamentals by their geometric mean.[1]

In this paper I show that the way the normalization enters the inversion algorithm matters for convergence. I compare the reference-fixed iteration with a *symmetric* iteration. I call it symmetric because every alternative is treated in the same way: the update is applied to every equation before a common normalization is imposed.

To clarify the comparison, let $\delta = (\delta_0, \ldots, \delta_J)'$ denote the fundamentals, $\hat{s}_j$ denote the observed target share of alternative j, and $\mathcal{S}_j(\delta)$ denote its model-implied counterpart. Define the log residual as

$$r_j(\delta) \equiv \log \hat{s}_j - \log \mathcal{S}_j(\delta).$$

If alternative q is used as the reference and normalized to $\delta_q = 0$, the reference-fixed update is

$$\delta_j^{t+1} = \delta_j^t + r_j(\delta^t), \qquad j \neq q, \tag{1}$$

while the symmetric update, with an *ex-post* normalization $\delta_q = 0$, is

$$\delta_j^{t+1} = \delta_j^t + r_j(\delta^t) - r_q(\delta^t), \qquad j = 0, \ldots, J. \tag{2}$$

[1] A prominent example is Ahlfeldt, Redding, Sturm, and Wolf (2015). In their Technical Appendix (pp. 40–41), they solve a BLP-type fixed-point problem in levels and express the fundamentals relative to their geometric mean.

These are different algorithms away from the fixed point. The reference-fixed iteration drops equation q, while the symmetric iteration uses it and normalizes only afterward. Setting $\delta_q = 0$ does not make the algorithms equivalent.

I establish three results. First, both algorithms converge globally from any finite starting value. The argument applies a theorem by Bifulco, Glück, Krebs, and Kukharskyy (2025) that upgrades local asymptotic stability of a fixed point to global attractiveness for non-expansive maps. I show that both the reference-fixed and symmetric maps satisfy the conditions for this theorem. The result provides, in particular, a formal convergence guarantee for the standard BLP iteration used in practice and for the corresponding ex-post-normalized inversion in the class of gravity models studied here. This corrects the common misconception that BLP proved (1) is a contraction mapping, as I discuss in the literature review below.

Second, the symmetric algorithm is invariant to the normalization. Fixing one alternative, imposing mean-zero log fundamentals, normalizing the geometric mean of fundamentals to one, or normalizing their sum in levels all generate the same sequence in relative fundamentals. The normalization is therefore only a choice of coordinates for the symmetric iteration. This is not true for the reference-fixed algorithm: changing the reference alternative changes which equation is removed and hence changes the iteration itself.

Third, I show theoretically that the symmetric iteration has a weakly smaller local convergence factor than the reference-fixed iteration for every choice of reference alternative. In other words, the symmetric iteration is locally weakly faster. The reference-fixed factor is bounded below by one minus the reference alternative's share, so a small reference share necessarily makes convergence slow. In standard BLP, the reference alternative is the outside good, so the bound explains why Dubé, Fox, and Su (2012, Table III) find slower convergence when the outside share is small. The symmetric iteration does not share this pathology. Although heterogeneity in individual probabilities can slow both algorithms, the symmetric iteration removes the dependence on the arbitrary choice of a reference alternative.

I show numerically that the theoretical speed gains are quantitatively important. For example, using the most demanding design in Pál and Sándor (2023), I find that the symmetric iteration reduces inner-loop iterations by 98.5 percent and the runtime from 36 minutes to less than 5 minutes, an 87 percent reduction.[2] Also, using the standard benchmarks of Nevo (2000)

[2] I parallelized across 10 cores using a computer with Windows 11 Pro, an Intel Core i9-10900K 3.70 GHz and

cereal data and the original BLP automobile data, as implemented by Conlon and Gortmaker (2020), I find that the symmetric iteration reduces market-share calculations by 55 and 11 percent, and runtime by 46 and 9 percent, respectively. The smaller gain in BLP is consistent with its much larger outside-good share, about 90 percent versus 52 percent in Nevo. Because the symmetric update requires only minor code changes, it should be the default procedure.

Related literature. As mentioned above, there is a common misconception regarding the original BLP convergence result. Berry, Levinsohn, and Pakes (1995) state in their main text that the standard fixed-point map is a contraction. This is not entirely correct. Their Appendix only proves contraction for a component-wise truncated operator on a bounded set. This establishes convergence of the truncated iteration and, together with their argument ruling out other fixed points, existence and uniqueness of the solution to the original unbounded system. It does not, however, establish convergence of the standard unbounded iteration that is actually used in practice. The distinction has often been overlooked in the subsequent literature.[3] I am the first to close this gap by proving global convergence of the standard iteration.

Similarly, quantitative trade and spatial papers use fixed-point inversion algorithms, including the class of gravity models in Allen, Arkolakis, and Takahashi (2020), but I am not aware of a formal global convergence proof for the ex-post-normalized inversion studied here.[4] I am the first to make this connection to Bifulco et al. (2025) explicit for these inversion algorithms.

Related work studies other computational approaches to demand estimation. Dubé et al. (2012) analyze the numerical performance of the nested fixed-point estimator and propose MPEC; Li (2018) formulates demand inversion as a convex minimization problem; Salanié and

64 GB of RAM.

[3]Without claiming to be exhaustive, prominent examples include Conlon and Gortmaker (2020, p. 1119), who write that BLP showed the standard iteration is a contraction mapping; Train (2009, pp. 322–323), who state that BLP proved that iterating the standard map converges to the fixed point; Nevo (2000, pp. 532–533), who describe the standard iteration as a "contraction mapping" and refer to BLP for a proof of convergence; and Dubé et al. (2012), who refer to "BLP's nested contraction-mapping algorithm." In a related nested-logit setting, the Appendix of Grigolon and Verboven (2014) applies BLP-style contraction conditions without the truncation used in BLP's proof. An exception is Fukasawa (2025, Remark 7), who also notes that the standard untruncated BLP operator is not, strictly speaking, a contraction on $\mathbb{R}^J$ and points to BLP's bounded-domain argument.

[4]Allen, Arkolakis, and Li (2024) provide a broader framework to study existence, uniqueness and convergence of equilibrium systems. Their contraction-based convergence result requires a strict global bound and therefore does not cover the boundary case in which the relevant global bound equals one, as it does for the inversion maps studied here. Their inversion examples A.5 and A.12 in their Online Appendix only establish up-to-a-constant uniqueness, but not convergence of an iterative algorithm.

Wolak (2022) propose a fast approximation to random-coefficients demand estimation; Conlon and Gortmaker (2020) discuss best practices; and Pál and Sándor (2023) and Monardo (2025) compare alternative algorithms. I show instead that the symmetric iteration is weakly faster than the reference-fixed iteration, without relying on approximations or constrained optimization methods such as MPEC.

A closely related paper is Fukasawa (2025). He proposes the same symmetric iteration as in (2), documents large speed gains over the standard BLP iteration (1), and derives global bounds that link convergence speed to heterogeneity. These bounds do not establish global convergence or provide a general ranking of the two iterations. Instead, I prove global convergence for both iterations and identify the mechanisms behind the speed gains. In particular, I derive their exact local convergence factors, separate the reference-share effect from heterogeneity, and show that the symmetric iteration's factor is weakly smaller for every choice of reference alternative. Finally, I relate the symmetric iteration to the broader class of normalizations used in quantitative trade, spatial, and urban models.

The choice of normalization has evolved differently across fields, apparently because of path dependence. The quantitative trade, spatial, and urban literatures stumbled upon the symmetric algorithm without recognizing it as a best practice. In contrast, demand estimation inherited from Berry (1994) and Berry et al. (1995) the apparently innocuous convention of fixing the outside good. Only later did work such as Dubé et al. (2012) study and document the resulting slow convergence when outside-good shares are small. By bringing these approaches into a common framework, my paper clarifies an inversion problem that is shared across fields.

2 A common inversion problem

There are $J+1$ alternatives, indexed by $j = 0, \ldots, J$, which could be destination locations, or products, and I types, indexed by i, which could represent origin locations, or consumer types. Type i has weight $\omega_i > 0$, with $\sum_i \omega_i = 1$. Conditional choice probabilities are

$$\pi_{ij}(\delta) = \frac{\exp(\delta_j + \mu_{ij})}{\sum_{k=0}^{J} \exp(\delta_k + \mu_{ik})}, \tag{3}$$

where δ_j is the fundamental to be recovered and μ_{ij} collects type-specific tastes or bilateral frictions. Aggregating over source types gives

$$\mathcal{S}_j(\delta) = \sum_i \omega_i \pi_{ij}(\delta).$$

The inversion problem is to recover δ such that

$$\mathcal{S}_j(\delta) = \hat{s}_j, \qquad j = 0, \ldots, J, \tag{4}$$

where $\hat{s}_j > 0$ and $\sum_j \hat{s}_j = 1$ are observed target shares.

Non-available alternatives. To include the possibility of non-available alternatives, I allow for $\mu_{ij} = -\infty$. I say that alternative j is available to type i when $\mu_{ij} > -\infty$. I assume that every type has at least one available alternative and that the alternatives are connected: for any j and m, there is a sequence $j = j_0, \ldots, j_L = m$ such that, for each $\ell = 1, \ldots, L$, both $j_{\ell-1}$ and j_ℓ are available to some type. The first assumption ensures that the choice probabilities are well-defined. Both assumptions are automatic when every $\mu_{ij} > -\infty$.[5]

The following are two examples of this common inversion problem.

Example 1: Random-coefficients demand. Here, i indexes consumers, j products, δ_j is mean utility, and μ_{ij} captures heterogeneous tastes. Aggregating over the distribution of heterogeneous tastes yields the market share function:

$$\mathcal{S}_j(\delta; \theta) = \int \frac{\exp(\delta_j + \mu_{ij})}{\sum_{k=0}^{J} \exp(\delta_k + \mu_{ik})} dF(\boldsymbol{\mu}_i; \theta),$$

where $F(\boldsymbol{\mu}_i; \theta)$ denotes the distribution of consumer heterogeneity, parameterized by θ. In practice, this integral is evaluated by simulation, and its discrete version takes the general form (3). The weights ω_i are the population weights of the consumer types or, with simulation, the weights attached to each draw; with equally weighted draws, $\omega_i = 1/I$. The targets $\hat{s}_j$ are ob-

[5]These assumptions are implied by Assumption 1 of Allen et al. (2020).

served market shares, so (4) is the usual BLP demand inversion of Berry (1994) and Berry et al. (1995).

Example 2: Gravity trade and output. In a gravity trade model, like the ones considered by Allen et al. (2020), let E_i denote expenditure in importing location i and let $\pi_{ij}(\delta)$ be the share of that expenditure allocated to goods from producing location j. Producer j's output satisfies

$$Y_j = \sum_i E_i \pi_{ij}(\delta).$$

Defining $\omega_i = E_i / \sum_\ell E_\ell$ and $\hat{s}_j = Y_j / \sum_k Y_k$ gives exactly (4). In this case, μ_{ij} collects bilateral trade frictions and δ_j contains the producer-side fundamental being recovered. A similar representation with commuting or migration flows can be obtained in urban and spatial models; see Ahlfeldt et al. (2015) and Caliendo, Dvorkin, and Parro (2019). Here, the possibility of $\mu_{ij} = -\infty$ is relevant: it means that importer i cannot buy from producer j, as may be the case for some pairs (i, j) in the data.

Up-to-a-constant identification. Because the model-implied shares are invariant to additive constants, meaning

$$\mathcal{S}(\delta + c\mathbf{1}) = \mathcal{S}(\delta),$$

the vector δ is identified only up to a common additive constant. Thus, a normalization is required.

Existence. I take existence of an interior solution as given. Existence is established for the demand applications by Berry (1994) and Berry et al. (1995), and for a broad classes of gravity models by Allen et al. (2020). Uniqueness will follow from the global-convergence result below.

2.1 Reference-fixed and symmetric iterations

Choose an alternative q and impose $\delta_q = 0$. The reference-fixed iteration drops equation q from the update and sets

$$T^R_{q,j}(\delta_{-q}) = \delta_j + \log \hat{s}_j - \log \mathcal{S}_j(\delta), \qquad j \neq q. \tag{5}$$

For BLP, $q = 0$ is the outside option and (5) is the standard fixed-point iteration.

In contrast, the symmetric iteration first updates every equation:

$$F_j(\delta) = \delta_j + \log \hat{s}_j - \log \mathcal{S}_j(\delta), \qquad j = 0, \ldots, J. \tag{6}$$

Because $F(\delta + c\mathbf{1}) = F(\delta) + c\mathbf{1}$, the update is then followed by a normalization. Let

$$N(\mathbf{x}) = \mathbf{x} - a(\mathbf{x})\mathbf{1}, \qquad a(\mathbf{x} + c\mathbf{1}) = a(\mathbf{x}) + c. \tag{7}$$

This class includes, for example, fixing any coordinate, mean-zero logs (which implies a geometric mean of one in levels), or sum-to-one in levels. The symmetric iteration is

$$\delta^{t+1} = N\big(F(\delta^t)\big)\,. \tag{8}$$

If N fixes coordinate q, the symmetric iteration takes the normalized form described in the Introduction. In the BLP application with $q = 0$, this is the $\gamma = 1$ update proposed by Fukasawa (2025).

While both algorithms are irrelevant for identification, their convergence behavior is different. The next sections discusses the speed advantages of the symmetric algorithm, both in theory and practice.

3 Global convergence, normalization, and speed

Here I present the main results. I first establish normalization invariance and global convergence, then compare convergence rates. All proofs are in the Appendix.

Proposition 1 (Normalization invariance)**.** *Consider two implementations of the symmetric iteration that use different normalizations. If their initial vectors differ only by a common constant, then their vectors differ only by a common constant at every subsequent iteration. Therefore they generate exactly the same path of relative fundamentals and have the same convergence rate in relative fundamentals.*

Thus, fixing one alternative after applying the full update is no different from imposing a geometric-mean or aggregate normalization. The normalization only chooses the common level at which the same relative fundamentals are reported. What matters is whether the normalized equation is removed from the update. The reference-fixed algorithm does this; the symmetric algorithm does not.

Proposition 2 (Global convergence)**.** *For any reference alternative q:*

(i) The reference-fixed iteration $\delta_{-q}^{t+1} = T_q^R(\delta_{-q}^t)$ converges to the unique solution normalized by $\delta_q = 0$ from any finite starting value.

(ii) The symmetric iteration (8) converges to the unique solution under any continuous normalization N satisfying (7), from any finite starting value.

The proof applies a result by Bifulco et al. (2025): on a connected metric space, a non-expansive map with a locally asymptotically stable fixed point is globally attractive. Both maps are globally non-expansive, and each Jacobian evaluated at the fixed point has spectral radius below one, which implies a locally asymptotically stable fixed point. The theorem therefore gives global convergence; Appendix B provides details.

Remark 1 (The standard BLP iteration)**.** *Part (i) of Proposition 2 includes the standard BLP iteration discussed in the Introduction. On $\mathbb{R}^J$, this map is globally non-expansive under the sup norm, but its global Lipschitz constant is one. More generally, it is not a contraction under any norm-induced metric. Proposition 2 therefore provides a direct convergence argument without imposing bounds or truncating the map, as BLP do in their Appendix.*

Convergence speed. I next compare the local linear convergence rates of the two algorithms. Evaluate probabilities at the solution δ^* and define

$$D \equiv \operatorname{diag}(\hat{\mathbf{s}}), \qquad M \equiv \sum_i \omega_i \boldsymbol{\pi}_i \boldsymbol{\pi}_i',$$

where $\boldsymbol{\pi}_i = (\pi_{i0}, \ldots, \pi_{iJ})'$. The Jacobian of the full symmetric update at the solution is

$$A \equiv \nabla F(\delta^*) = D^{-1} M. \tag{9}$$

The matrix A is row stochastic, and by Lemma 1, irreducible. Moreover,

$$B \equiv D^{1/2} A D^{-1/2} = D^{-1/2} M D^{-1/2} \tag{10}$$

is symmetric and positive semidefinite. Hence the eigenvalues of A can be ordered as

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{J+1} \geq 0.$$

Proposition 3 (Local convergence rates)**.** *Suppose the normalization N is differentiable at the solution. Let ρ^S denote the asymptotic local linear convergence factor of the symmetric iteration, and ρ_q^R denote the factor of the reference-fixed iteration that removes alternative q. For a matrix X, let $\lambda_{\max}(X)$ denote its largest eigenvalue and $X_{-q,-q}$ its principal submatrix obtained by deleting its qth row and column. Then:*

(i) The symmetric rate is

$$\rho^S = \lambda_2. \tag{11}$$

(ii) The reference-fixed rate is

$$\rho_q^R = \lambda_{\max}\left(A_{-q,-q}\right),$$

and for every q,

$$\rho^S \leq \rho_q^R. \tag{12}$$

(iii) The reference-fixed rate satisfies

$$\rho_q^R \geq 1 - \hat{s}_q + \frac{\mathrm{Var}(\pi_{iq})}{1 - \hat{s}_q} \geq 1 - \hat{s}_q. \tag{13}$$

(iv) The symmetric rate can be written as

$$\rho^S = \lambda_{\max}\left(D^{-1/2}\,\mathrm{Cov}(\boldsymbol{\pi}_i) D^{-1/2}\right). \tag{14}$$

The key comparison in (12) follows from the eigenvalue interlacing theorem (Meyer, 2000, pp. 552–553). The reference-fixed Jacobian $A_{-q,-q}$ is similar to the principal submatrix $B_{-q,-q}$ of the symmetric matrix defined in (10). Its largest eigenvalue must therefore lie between $\lambda_1 = 1$ and λ_2. The symmetric algorithm removes the unit eigenvalue directly by working with relative fundamentals, while the reference-fixed algorithm inherits a principal submatrix whose largest eigenvalue is weakly larger than λ_2.

Equation (13) separates two sources that enter a lower bound on the convergence rate of the reference-fixed iteration. A small share of the normalized alternative mechanically pushes the convergence factor toward one, and the bound also contains heterogeneity in the probability of choosing that alternative. This explains the familiar observation that the standard BLP iteration becomes slow when the outside share is small; see Dubé et al. (2012), Conlon and Gortmaker (2020), and Fukasawa (2025). The symmetric algorithm does not depend on an arbitrary reference-share term. Equation (14) shows instead that its remaining convergence rate is governed by heterogeneity in individual choice probabilities, standardized by the observed shares.

No heterogeneity. If $\boldsymbol{\pi}_i = \hat{\mathbf{s}}$ for every type, we have the standard logit model. Then $\mathrm{Cov}(\boldsymbol{\pi}_i) = 0$ and

$$\rho^S = 0.$$

The symmetric iteration recovers relative fundamentals in one step.[6] In contrast, the reference-fixed Jacobian has rank one and

$$\rho^R_q = 1 - \hat{s}_q.$$

Thus, even in the logit model, a reference-fixed iteration can be arbitrarily slow when the reference share is small.

One nonreference alternative. With two alternatives in total, or one "inside" good in BLP lingo, the rates have particularly simple expressions. Let q be the reference alternative. Then

$$\rho^S = \frac{\mathrm{Var}(\pi_{iq})}{\hat{s}_q(1-\hat{s}_q)}, \qquad \rho^R_q = 1 - \hat{s}_q + \frac{\mathrm{Var}(\pi_{iq})}{1-\hat{s}_q}.$$

[6]This was already noted by Fukasawa (2025).

Since $\text{Var}(\pi_{iq}) \leq \hat{s}_q(1-\hat{s}_q)$, the symmetric rate is always weakly smaller, with strict inequality for finite logit probabilities. As heterogeneity approaches deterministic sorting, so that each consumer type effectively chooses a single product, both rates approach one and the advantage of symmetric normalization becomes small.

There is no analogous nontrivial bound on ρ^S using aggregate shares alone. For any interior vector $\hat{\mathbf{s}}$, the symmetric rate is zero when all types have the same choice probabilities and can be arbitrarily close to one when heterogeneity increases and types sort almost deterministically across alternatives. This difference is useful empirically: the observed share of the reference alternative alone can diagnose a source of slow convergence for the reference-fixed algorithm, while the speed of the symmetric algorithm depends on the underlying heterogeneity.

4 Numerical exercises

So far, I have shown that the symmetric iteration is weakly faster than the reference-fixed iteration. The numerical exercises of this section ask whether the local ranking in Proposition 3 is quantitatively important.[7] Every comparison uses the same simulated consumers and starting values; only the inversion map changes. To make the stopping criterion comparable across methods, define the distance between observed and predicted shares as

$$d\big(\hat{\mathbf{s}}, \mathcal{S}(\delta)\big) = \max_{j \neq 0} \left|\log \hat{s}_j - \log \mathcal{S}_j(\delta)\right|.$$

The outside share equation is redundant and is omitted for both methods. Both algorithms stop when this distance is at most 10^{-12}. One market-share calculation means one evaluation of $\mathcal{S}(\delta)$. The symmetric implementation updates all alternatives and then sets the outside-good utility to zero, as in (2). To isolate the effect of this change, I use a basic iteration throughout, with no accelerations of any kind.

I first use a fixed random-coefficients demand environment in the spirit of Dubé et al. (2012). There are $J = 25$ inside goods and one outside good, and market shares are computed using $I = 1{,}000$ equally weighted taste draws. The same product realization and taste draws are held

[7] All the code to replicate this section can be found in https://github.com/zerecero/norm_conv_speed

fixed throughout. One experiment varies a common intercept on inside-good mean utilities, moving the outside share while holding the remaining environment fixed. A second scales preference heterogeneity and re-centers the intercept to keep the outside share at 0.10. All the details of the numerical exercises are in the Supplemental Material. Figure 1 reports market-share calculations and the local convergence factors.

Figure 1 shows large differences. As the outside share falls, the reference-fixed rate approaches one and its share calculations rise sharply, while the symmetric rate remains lower. Holding the outside share fixed, heterogeneity slows both algorithms, but the symmetric iteration remains faster. The Figure also shows that variance-adjusted bound closely tracks the reference-fixed rate. The simpler $1 - \hat{s}_0$ bound is also close.

Next, I consider the comparatively difficult full-estimation design from Section 7 of Dubé et al. (2012), following Pál and Sándor (2023) and Monardo (2025). It has $T = 50$ markets, $J = 25$ inside goods, and intercepts $\{-4, -2, 0, 2, 4\}$, moving the mean outside share from 0.94 to 0.34. I simulate 20 datasets per intercept. For each estimation, I do five parameter starts. The estimation uses PyBLP, 1,331 quasi-random draws, and ten-process market-level parallelization. Only the inversion map changes. I put a cap of 70,000 inner iterations. If the algorithm hits this cap it will be recorded as not converging. Table 1 reports convergence over all runs; runtime and mean inner iterations use replications in which all five starts converge under both algorithms.

The methods perform similarly when the outside share is high. As it falls, reference-fixed performance deteriorates sharply while symmetric performance remains stable. At intercept 4, reference-fixed convergence is 62 percent, compared with 100 percent for the symmetric map. In the paired-complete sample, mean iterations per market inversion fall from 1,499.9 to 22.2, a 98.5 percent reduction, and runtime for five starts falls from 36.3 to 4.6 minutes, an 87.5 percent reduction. Symmetric runtime remains close to five minutes throughout. Because capped reference-fixed executions are omitted from the speed columns, the comparison understates its disadvantage in the hardest designs. Thus, the mechanism isolated in Figure 1 remains consequential inside a full estimation.

To see if this matters in practice, I re-estimate two common benchmarks in the literature: the standard Nevo (2000) cereal model and the original BLP automobile model as implemented in Conlon and Gortmaker (2020). Table 2 changes only the inversion map. In Nevo, both methods

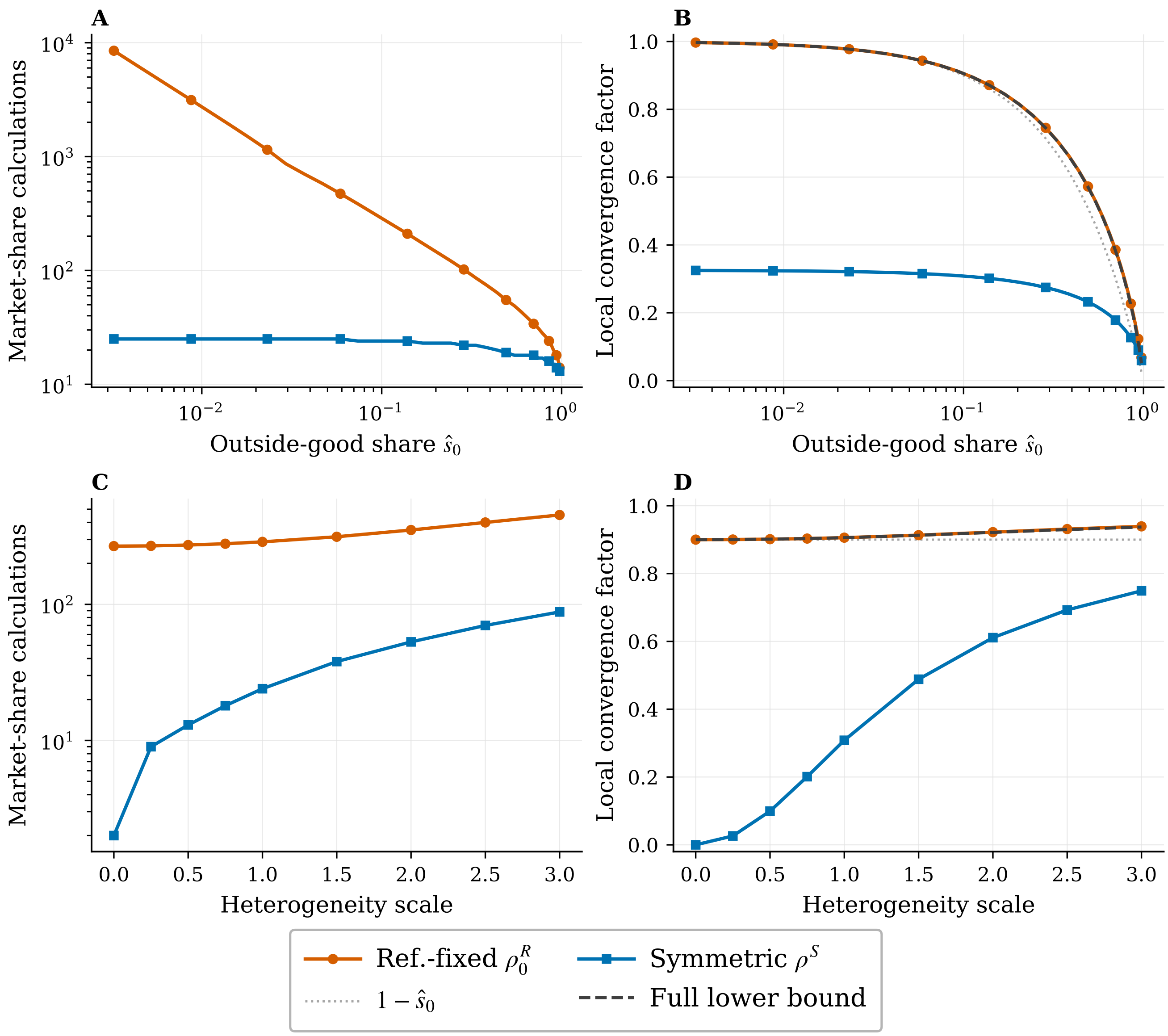

Figure 1: Normalization and convergence in controlled random-coefficients demand inversions. Panels A and B vary a common inside-good intercept and report results against the induced outside-good share. Panels C and D vary preference heterogeneity while holding the outside-good share at 0.10. The left panels report market-share calculations. The right panels report the local convergence factors, $1 - \hat{s}_0$, and the variance-adjusted lower bound in (13).

converge in 57 objective evaluations to the same objective value. The symmetric map reduces market-share calculations from 637,718 to 283,828, a 55.5 percent reduction, and runtime falls from 63.3 to 34.3 seconds. In original BLP, both methods converge in 66 objective evaluations to the same objective value. Market-share calculations fall from 401,818 to 356,930, an 11.2 percent reduction, and runtime falls from 140.2 to 127.3 seconds.

5 Conclusion

This paper studies two ways of incorporating normalization into a common class of demand and gravity inversions. A reference-fixed iteration normalizes one alternative and removes its equation from the update. A symmetric iteration updates all equations and normalizes afterwards. I find that both procedures converge globally, but the symmetric iteration weakly dominates every reference-fixed iteration in local convergence speed.

I show that reference-fixed convergence is necessarily slow when the removed alternative has a small share, and its lower bound also contains a heterogeneity term. This gives a clear theoretical explanation to the empirical patterns documented in the literature. The symmetric algorithm removes the dependence on the arbitrary reference share; its remaining rate reflects heterogeneity in the underlying choice probabilities.

The numerical exercises show that this distinction is quantitatively important. Using standard benchmarks, the symmetric map substantially reduces function evaluations and cuts both inner-loop work and runtime. Since it requires the same share calculation as the reference-fixed update and only a different normalization step, the reference-fixed offers no practical advantage. The same lesson carries over to quantitative trade, spatial, and urban models, where estimation often nests repeated inversions inside an outer loop. For this class of inversion problems, symmetric normalization should remain the default.

Table 1: Full estimation in a difficult Monte Carlo design.

Intercept	Mean $\hat{s}_0$	Converged (%) Ref.	Converged (%) Sym.	Paired reps.	Running time (min) Ref.	Running time (min) Sym.	Mean inner iterations Ref.	Mean inner iterations Sym.
-4	0.9373	100	100	20	5.39	5.34	26.5	16.4
-2	0.8619	100	100	20	5.78	5.36	50.0	19.5
0	0.7336	100	100	20	7.73	5.01	136.1	19.1
$+2$	0.5520	90	100	18	16.34	4.73	455.0	20.3
$+4$	0.3428	62	100	11	36.33	4.55	1499.9	22.2

Notes: The design follows Section 7 of Dubé et al. (2012), as implemented by Pál and Sándor (2023) and Monardo (2025). There are 20 replications and five starts per replication, giving 100 executions per algorithm and intercept. Mean $\hat{s}_0$ averages the outside share across markets and replications. Paired reps. is the number of replications in which all five starts converge under both algorithms. Running time is the total wall-clock time, in minutes, required for the five sequential starts, averaged over these paired-complete replications. Mean inner iterations are total iterations divided by market inversions over the same sample. Markets were parallelized over ten processes. Both algorithms use identical PyBLP data, integration draws, instruments, starts, outer optimization, and a 10^{-12} inversion tolerance; only the inversion map changes. Each market inversion is capped at 70,000 iterations. Capped executions count against convergence but are excluded from the conditional speed columns. Computations were performed on Windows 11 Pro using a 10-core Intel Core i9-10900K 3.70 GHz processor with 64 GB of RAM.

Table 2: Full estimation benchmarks with Nevo and BLP data.

Dataset	Mean $\hat{s}_0$	Algorithm	Obj. evals.	Share calcs.	Time (s)	Objective	Reduction (%)
Nevo (2001)	0.524	Reference-fixed	57	637,718	63.3	4.5615	—
		Symmetric	57	283,828	34.3	4.5615	55.5
BLP (1995)	0.892	Reference-fixed	66	401,818	140.2	497.3357	—
		Symmetric	66	356,930	127.3	497.3357	11.2

Notes: Estimates use PyBLP 1.2.0 and its packaged data. Nevo uses the standard demand specification with one GMM step; BLP uses the standard demand-and-supply specification with two GMM steps. Within each dataset, the data, starting values, optimizer, and stopping rule are held fixed; only the unaccelerated inversion map changes. Inner solves stop when $\max_{j \neq 0} |\log \hat{s}_j - \log \mathcal{S}_j(\delta)| \leq 10^{-12}$. Share calcs. is the total number of market-share function evaluations across inner solves. Mean $\hat{s}_0$ is the simple average observed outside-good share across markets. Reduction is the percent decrease in share calculations. Times are from serial runs. Computations were performed on Windows 11 Pro using a 10-core Intel Core i9-10900K 3.70 GHz processor with 64 GB of RAM.

References

AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

ALLEN, T., C. ARKOLAKIS, AND X. LI (2024): "On the equilibrium properties of spatial models," *American Economic Review: Insights*, 6, 472–489.

ALLEN, T., C. ARKOLAKIS, AND Y. TAKAHASHI (2020): "Universal gravity," *Journal of Political Economy*, 128, 393–433.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

BERRY, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

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.

——— (2026): "Single and attractive: Uniqueness and stability of economic equilibria under monotonicity assumptions," *Journal of International Economics*, 104187.

CALIENDO, L., M. DVORKIN, AND F. PARRO (2019): "Trade and labor market dynamics: General equilibrium analysis of the china trade shock," *Econometrica*, 87, 741–835.

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.

FUKASAWA, T. (2025): "Fast and simple inner-loop algorithms of static/dynamic BLP estimations," *arXiv preprint arXiv:2404.04494*.

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*.

MEYER, C. (2000): "Matrix Analysis and Applied Linear Algebra. Society for Industrial and Applied Mathematics (SIAM)," *PA, USA*.

MONARDO, J. (2025): "Comparing methods for estimating random coefficient logit demand models," Working paper, University of Bristol.

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.

PÁL, L. AND Z. SÁNDOR (2023): "Comparing procedures for estimating random coefficient logit demand models with a special focus on obtaining global optima," *International Journal of Industrial Organization*, 88, 102950.

SALANIÉ, B. AND F. A. WOLAK (2022): "Fast, detail-free, and approximately correct: Estimating mixed demand systems," Tech. rep., Working paper.

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

APPENDIX

A Proof of Proposition 1

Suppose two implementations use normalizations N and $\widetilde{N}$ and start from

$$\widetilde{\boldsymbol{\delta}}^0 = \boldsymbol{\delta}^0 + c_0\mathbf{1}$$

for some constant c_0. The full update satisfies

$$F(\boldsymbol{\delta} + c\mathbf{1}) = F(\boldsymbol{\delta}) + c\mathbf{1},$$

because a common shift leaves all model shares unchanged. Therefore

$$F(\widetilde{\boldsymbol{\delta}}^0) = F(\boldsymbol{\delta}^0) + c_0\mathbf{1}.$$

Each normalization only subtracts a common constant. Hence there is another constant c_1 such that

$$\widetilde{\boldsymbol{\delta}}^1 = \boldsymbol{\delta}^1 + c_1\mathbf{1}.$$

Repeating the same argument gives, for every t,

$$\widetilde{\boldsymbol{\delta}}^t = \boldsymbol{\delta}^t + c_t\mathbf{1}.$$

It follows that

$$\widetilde{\delta}^t_j - \widetilde{\delta}^t_k = \delta^t_j - \delta^t_k$$

for every pair j, k and every iteration t. The two implementations therefore generate exactly the same sequence of relative fundamentals, and hence the same convergence rate in those relative fundamentals.

B Proof of Proposition 2

Following Bifulco et al. (2025), let (Z, d) be a metric space, let $G : Z \to Z$, and let z^* be a fixed point of G. The space (Z, d) is *connected* if it cannot be written as the union of two nonempty, open, disjoint subsets. The map G is *globally non-expansive* if

$$d(G(z), G(z')) \leq d(z, z')$$

for all $z, z' \in Z$; equivalently, G is Lipschitz continuous with Lipschitz constant at most one. The fixed point z^* is *Lyapunov stable* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $d(z, z^*) < \delta$ implies $d(G^n(z), G^n(z^*)) < \varepsilon$ for all integers $n \geq 0$. It is *locally attractive* if there exists $\delta > 0$ such that $d(z, z^*) < \delta$ implies $G^n(z) \to z^*$ as $n \to \infty$. Finally, z^* is *locally asymptotically stable* if it is both Lyapunov stable and locally attractive, and *globally attractive* if $G^n(z) \to z^*$ for every $z \in Z$.

Now I present the global convergence theorem used later on.

Theorem 1 (Bifulco et al. (2025), Theorem C.1)**.** *Let (Z, d) be a connected metric space and let $G : Z \to Z$ be Lipschitz continuous with Lipschitz constant at most one. If z^* is a locally asymptotically stable fixed point of G, then z^* is globally attractive and unique.*

The proof of Theorem 1 is in the Supplemental Material. This was originally an auxiliary result for their main theorem concerning uniqueness and attractivity of equilibria. Their published paper Bifulco, Glück, Krebs, and Kukharskyy (2026) uses a different strategy.

The next Lemma presents a useful implication of connected alternatives.

Lemma 1. *If the alternatives are connected, the matrix $A \equiv \nabla F(\delta^*)$ is irreducible.*

Proof. Differentiating the full map (6) gives, for every j, m,

$$\frac{\partial F_j}{\partial \delta_m} = \frac{\sum_i \omega_i \pi_{ij}(\delta) \pi_{im}(\delta)}{\mathcal{S}_j(\delta)} \geq 0. \tag{15}$$

For every j and m, (15) implies

$$A_{jm} > 0 \quad \iff \quad j \text{ and } m \text{ are available to some type.}$$

Indeed, all terms in the numerator of (15) are nonnegative, and $\pi_{ij}(\delta) > 0$ if and only if $\mu_{ij} > -\infty$, for any δ, including the fixed-point.

For any j and m, connectedness gives a sequence $j = j_0, \ldots, j_L = m$ such that

$$A_{j_{\ell-1} j_\ell} > 0, \qquad \ell = 1, \ldots, L.$$

Because $A \geq 0$, the product along this path is one of the nonnegative terms in the expansion of $(A^L)_{jm}$. Hence

$$(A^L)_{jm} \geq \prod_{\ell=1}^{L} A_{j_{\ell-1} j_\ell} > 0.$$

Thus, for every distinct j and m, some power of A has a positive (j, m) entry. Hence A is irreducible. $\square$

B.1 Reference-fixed iteration

For $j, m \neq q$, differentiating (5) gives

$$\frac{\partial T^R_{q,j}}{\partial \delta_m} = \frac{\sum_i \omega_i \pi_{ij}(\delta)\pi_{im}(\delta)}{\mathcal{S}_j(\delta)} \geq 0. \tag{16}$$

The jth row sum is

$$\sum_{m \neq q} \frac{\partial T^R_{q,j}}{\partial \delta_m} = 1 - \frac{\sum_i \omega_i \pi_{ij}(\delta)\pi_{iq}(\delta)}{\mathcal{S}_j(\delta)} \leq 1 \tag{17}$$

for every finite δ; equality holds if j and q are not available to any common type. Hence

$$\|\nabla T^R_q(\delta)\|_\infty \leq 1$$

The mean-value inequality therefore implies that T^R_q is globally non-expansive under the sup norm. Non-expansiveness immediately implies Lyapunov stability of the fixed point.

It remains to verify local attractiveness. Let $\rho(X)$ denote X's spectral radius. Without loss of generality, relabel q as the last alternative and define

$$C_q \equiv \begin{pmatrix} A_{-q,-q} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix}.$$

The matrix C_q has the same dimension as A. We have that $\rho(A_{-q,-q}) = \rho(C_q)$. Since $A\mathbf{1} = \mathbf{1}$ and A is nonnegative and irreducible, the Perron–Frobenius theorem (Meyer, 2000, p. 673) gives $\rho(A) = 1$. Wielandt's theorem (Meyer, 2000, p. 675) gives $\rho(C_q) \leq \rho(A)$, with equality only if $|C_q| = A$. But $|C_q| = C_q \neq A$, because C_q has a zero row and column whereas A is irreducible. Therefore,

$$\rho(A_{-q,-q}) = \rho(C_q) < \rho(A) = 1.$$

By the Ostrowski local attraction theorem (Ortega and Rheinboldt, 1970, Theorem 10.1.3), the fixed point is locally attractive. As it is also Lyapunov stable, it is therefore locally asymptotically stable. Since $\mathbb{R}^J$ is connected, Theorem 1 gives global convergence and uniqueness.

For completeness, the untruncated reference-fixed map is generally not a contraction on the whole space. Consider $\delta_j^{(M)} = \delta_j + M$ for all $j \neq q$, with $\delta_q = 0$. Hence, for every $j \neq q$, $\pi_{ij}(\boldsymbol{\delta}^{(M)})\pi_{iq}(\boldsymbol{\delta}^{(M)}) \to 0$ for every i, while $\mathcal{S}_j(\boldsymbol{\delta}^{(M)})$ converges to a positive value. Then, every row sum in (17) converges to one. By the Collatz–Wielandt inequality (see Meyer (2000, pp. 666–668)),

$$\rho\left(\nabla T_q^R(\boldsymbol{\delta}^{(M)})\right) \to 1.$$

As every induced matrix norm dominates the spectral radius (Horn and Johnson, 2013, Theorem 5.6.9), the global Lipschitz constant is at least one under every norm-induced metric. This is the issue behind Remark 1.

B.2 Symmetric iteration

By Proposition 1, it is enough to prove convergence under one convenient normalization. Take

$$N_w(\mathbf{x}) = \mathbf{x} - (w'\mathbf{x} - c)\mathbf{1}, \qquad w_j \geq 0, \qquad w'\mathbf{1} = 1,$$

and let

$$X_w = \{\mathbf{x} : w'\mathbf{x} = c\}.$$

This includes fixing a coordinate and, with equal weights, imposing a geometric mean of one in levels. The normalized map $G_w = N_w \circ F$ maps X_w into itself. Define on X_w

$$d_Q(\mathbf{x}, \mathbf{y}) \equiv \inf_{a \in \mathbb{R}} \|\mathbf{x} - \mathbf{y} - a\mathbf{1}\|_\infty.$$ [8]

Since the normalization fixes the common level, d_Q is a metric on the connected set X_w.

The full map F is globally non-expansive under the sup norm because its Jacobian (15) is nonnegative and every row sums to one. Using shift equivariance, $F(\mathbf{y} + a\mathbf{1}) = F(\mathbf{y}) + a\mathbf{1}$,

$$\begin{aligned} d_Q(F(\mathbf{x}), F(\mathbf{y})) &= \inf_a \|F(\mathbf{x}) - F(\mathbf{y} + a\mathbf{1})\|_\infty \\ &\leq \inf_a \|\mathbf{x} - \mathbf{y} - a\mathbf{1}\|_\infty \\ &= d_Q(\mathbf{x}, \mathbf{y}). \end{aligned}$$

[8] Equivalently, $d_Q(\mathbf{x}, \mathbf{y}) = \frac{1}{2}[\max_j(x_j - y_j) - \min_j(x_j - y_j)]$. If $M = \max_j(x_j - y_j)$ and $m = \min_j(x_j - y_j)$, the minimizing shift is $a = (M + m)/2$. Thus d_Q ignores common additive shifts. For positive level vectors obtained by exponentiating $\mathbf{x}$ and $\mathbf{y}$, Hilbert's projective metric is $2d_Q$.

Since N_w only subtracts a common constant,

$$d_Q(G_w(\mathbf{x}), G_w(\mathbf{y})) = d_Q(F(\mathbf{x}), F(\mathbf{y})),$$

so G_w is globally non-expansive on (X_w, d_Q), which implies the fixed-point δ^* is Lyapunov stable.

It remains to establish that δ^* is locally attractive. At the fixed point, let A be defined by (9) and set

$$P_w = I - \mathbf{1}w'.$$

Then

$$\nabla G_w(\delta^*) = P_w A.$$

The similarity transformation in (10) is symmetric positive semidefinite. Since A is nonnegative and irreducible, and $A\mathbf{1} = \mathbf{1}$, the Perron–Frobenius theorem (Meyer, 2000, p. 673) and (10) imply that its eigenvalues satisfy

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{J+1} \geq 0.$$

As A is diagonalizable, the corresponding eigenvectors $\mathbf{v}_1 = \mathbf{1}, \mathbf{v}_2, \ldots, \mathbf{v}_{J+1}$ are linearly independent. Normalization replaces the first eigenvalue by zero and leaves the others unchanged. Indeed, $A\mathbf{1} = \mathbf{1}$ implies

$$P_w A\mathbf{1} = 0.$$

For $k = 2, \ldots, J+1$, define $\widetilde{\mathbf{v}}_k = P_w \mathbf{v}_k$. Since $P_w \mathbf{1} = 0$,

$$\begin{aligned} P_w A \widetilde{\mathbf{v}}_k &= P_w A \left(\mathbf{v}_k - (w'\mathbf{v}_k)\mathbf{1}\right) \\ &= \lambda_k P_w \mathbf{v}_k = \lambda_k \widetilde{\mathbf{v}}_k. \end{aligned}$$

The vectors $\widetilde{\mathbf{v}}_2, \ldots, \widetilde{\mathbf{v}}_{J+1}$ are linearly independent. Otherwise, a linear combination of the original eigenvectors $\mathbf{v}_2, \ldots, \mathbf{v}_{J+1}$ would be a multiple of $\mathbf{1}$, contradicting their linear independence. Therefore

$$\operatorname{spec}(P_w A) = \{0, \lambda_2, \ldots, \lambda_{J+1}\},$$

and

$$\rho(P_w A) = \lambda_2 < 1.$$

By the Ostrowski local attraction theorem (Ortega and Rheinboldt, 1970, Theorem 10.1.3), δ^* is locally attractive in $(\mathbb{R}^{J+1}, \|\cdot\|_\infty)$. On X_w,

$$\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_\infty \le d_Q(\mathbf{x},\mathbf{y}) \le \|\mathbf{x}-\mathbf{y}\|_\infty.$$ [9]

Hence the fixed-point δ^* is also locally attractive in (X_w, d_Q). Together with Lyapunov stability, δ^* is therefore locally asymptotically stable in (X_w, d_Q).

All the conditions of Theorem 1 now hold on (X_w, d_Q), so the fixed point is globally attractive and unique under this normalization. Proposition 1 then extends convergence to any continuous normalization satisfying (7), since all such normalizations generate the same relative fundamentals. This proves part (ii) of Proposition 2.

C Proof of Proposition 3

The asymptotic local convergence factor equals the spectral radius of its Jacobian at the fixed point (Ortega and Rheinboldt, 1970, Theorem 10.1.4, p. 301). What follows characterize them.

By the argument in Appendix B, the eigenvalues of A, the Jacobian of the full map evaluated at the fixed point, satisfy

$$1 = \lambda_1 > \lambda_2 \ge \cdots \ge \lambda_{J+1} \ge 0.$$

Write the normalization as in (7). At the solution, let

$$w = \nabla a(\delta^*).$$

The shift property in (7) implies $w'\mathbf{1} = 1$ and the normalization has Jacobian[10]

$$\nabla N(\delta^*) = I - \mathbf{1}w'.$$

[9]Let $\mathbf{h} = \mathbf{x}-\mathbf{y}$. Since $\mathbf{x},\mathbf{y} \in X_w$, $w'\mathbf{h} = 0$. The upper bound follows by setting $a = 0$ in the definition of d_Q. For any a, $w'(\mathbf{h} - a\mathbf{1}) = -a$. Since $w_j \ge 0$ and $w'\mathbf{1} = 1$, $|a| \le \sum_j w_j |h_j - a| \le \|\mathbf{h} - a\mathbf{1}\|_\infty$. Moreover, $\mathbf{h} = (\mathbf{h} - a\mathbf{1}) + a\mathbf{1}$, so the triangle inequality gives $\|\mathbf{h}\|_\infty \le \|\mathbf{h} - a\mathbf{1}\|_\infty + \|a\mathbf{1}\|_\infty = \|\mathbf{h} - a\mathbf{1}\|_\infty + |a| \le 2\|\mathbf{h} - a\mathbf{1}\|_\infty$. Taking the infimum over a gives the lower bound.

[10]Let $g(c) = a(\delta + c\mathbf{1})$. The shift property gives $g(c) = a(\delta) + c$, so $g'(0) = 1$. By the chain rule, $g'(0) = \nabla a(\delta)'\mathbf{1}$, and therefore, at the solution, $w'\mathbf{1} = 1$. Moreover, $N_j(\delta) = \delta_j - a(\delta)$, so $\partial N_j / \partial \delta_m = \mathbf{1}\{j = m\} - \partial a / \partial \delta_m$. Evaluating at δ^* gives $\nabla N(\delta^*) = I - \mathbf{1}w'$.

Thus the Jacobian of the normalized symmetric iteration is

$$(I - \mathbf{1}w')A.$$

The projection argument in Appendix B applies for any w satisfying $w'\mathbf{1} = 1$: it changes the spectrum from $\{1, \lambda_2, \ldots, \lambda_{J+1}\}$ to $\{0, \lambda_2, \ldots, \lambda_{J+1}\}$. Hence the asymptotic local convergence factor of the symmetric iteration is λ_2, proving (11).

For the reference-fixed iteration, (16) evaluated at the solution gives

$$\nabla T_q^R(\boldsymbol{\delta}^*) = A_{-q,-q}.$$

Moreover,

$$D_{-q}^{1/2} A_{-q,-q} D_{-q}^{-1/2} = B_{-q,-q},$$

so $A_{-q,-q}$ has the same eigenvalues as the principal submatrix $B_{-q,-q}$, which is positive semidefinite. Let its largest eigenvalue be $\mu_1^{(q)}$. By the eigenvalues interlacing theorem (Meyer, 2000, pp. 552–553),

$$\lambda_1 \geq \mu_1^{(q)} \geq \lambda_2.$$

Then,

$$\rho_q^R = \mu_1^{(q)} \geq \lambda_2 = \rho^S,$$

which proves (12).

To obtain the observable lower bound, apply the Rayleigh quotient (Meyer, 2000, pp. 549–550) to $B_{-q,-q}$ using the vector $\mathbf{x} = (\sqrt{\hat{s}_j})_{j \neq q}$. Then

$$\begin{aligned}
\rho_q^R \geq \frac{\mathbf{x}' B_{-q,-q} \mathbf{x}}{\mathbf{x}'\mathbf{x}} &= \frac{\sum_{j \neq q} \sum_{m \neq q} M_{jm}}{1 - \hat{s}_q} \\
&= \frac{\sum_{j \neq q} \sum_{m \neq q} \mathbb{E}[\pi_{ij}\pi_{im}]}{1 - \hat{s}_q} = \frac{\mathbb{E}\left[\left(\sum_{j \neq q} \pi_{ij}\right)\left(\sum_{m \neq q} \pi_{im}\right)\right]}{1 - \hat{s}_q} \\
&= \frac{\mathbb{E}\left[\left(\sum_{j \neq q} \pi_{ij}\right)^2\right]}{1 - \hat{s}_q} = \frac{\mathbb{E}\left[(1 - \pi_{iq})^2\right]}{1 - \hat{s}_q} \\
&= \frac{(1 - \hat{s}_q)^2 + \mathrm{Var}(\pi_{iq})}{1 - \hat{s}_q} = 1 - \hat{s}_q + \frac{\mathrm{Var}(\pi_{iq})}{1 - \hat{s}_q} \geq 1 - \hat{s}_q.
\end{aligned}$$

Here expectations and variances use the weights ω_i. The penultimate equality uses $\mathbb{E}[\pi_{iq}] = \hat{s}_q$

at the solution. This proves (13).

Finally, write

$$\Sigma_\pi \equiv \mathrm{Cov}(\boldsymbol{\pi}_i).$$

Since

$$\Sigma_\pi = \mathbb{E}(\boldsymbol{\pi}_i \boldsymbol{\pi}_i') - \mathbb{E}(\boldsymbol{\pi}_i)\mathbb{E}(\boldsymbol{\pi}_i)',$$

we have that

$$M = \hat{\mathbf{s}}\hat{\mathbf{s}}' + \Sigma_\pi,$$

and

$$B = \sqrt{\hat{\mathbf{s}}}\sqrt{\hat{\mathbf{s}}}' + D^{-1/2}\Sigma_\pi D^{-1/2}.$$

Because $\Sigma_\pi \mathbf{1} = 0$,

$$D^{-1/2}\Sigma_\pi D^{-1/2}\sqrt{\hat{\mathbf{s}}} = D^{-1/2}\Sigma_\pi \mathbf{1} = 0.$$

Also,

$$\sqrt{\hat{\mathbf{s}}}\sqrt{\hat{\mathbf{s}}}'\sqrt{\hat{\mathbf{s}}} = \sqrt{\hat{\mathbf{s}}}\sum_j \hat{s}_j = \sqrt{\hat{\mathbf{s}}}.$$

Thus the rank-one term $\sqrt{\hat{\mathbf{s}}}\sqrt{\hat{\mathbf{s}}}'$ generates the unit eigenvalue, while the remaining eigenvalues are those of $D^{-1/2}\Sigma_\pi D^{-1/2}$. Therefore

$$\rho^S = \lambda_2 = \lambda_{\max}\left(D^{-1/2}\Sigma_\pi D^{-1/2}\right),$$

which proves (14).