Combinatorial Discrete Choice: Theory and Application to Multinational Production

Costas Arkolakis, Fabian Eckert, and Rowan Shi*

March 29, 2026

In general, finding the optimal combination among N items requires evaluating all 2^N possibilities. We develop a method applicable whenever items are complements or substitutes that uses this structure to discard suboptimal combinations without evaluation and extends to heterogeneous-agent environments. In a calibrated general equilibrium model of multinational firms choosing interdependent production locations, we show the method is orders of magnitude faster than incumbent approaches and avoids error from discretizing firm heterogeneity. Quantitatively, whether production locations are complements or substitutes determines whether welfare gains from multinational production concentrate in the most productive countries or accrue more broadly.

*Arkolakis: Yale University (costas.arkolakis@yale.edu); Eckert: University of California, San Diego (fpe@ucsd.edu); Shi: Toronto Metropolitan University (rowanxshi@torontomu.ca). We are indebted to Junyuan Chen for help with the Julia code. We are grateful to Pablo Fajgelbaum and Eduardo Morales for excellent discussions at the Princeton IES Summer Trade Workshop. We thank Stanislav Atanasov, Jake DiCicco, Tra Nguyen, Andrew Salmon, Benjamin Tong, Yunus Tuncbilek, and Nardeen Abdulkareem for outstanding research assistance. We also thank Chrysanthos Gounaris, David Atkin, Arnaud Costinot, Dave Donaldson, Penny Goldberg, Navin Kartik, Samuel Kortum, Nitya Pandalai-Nayar, and Steve Redding for comments and suggestions. Shi gratefully acknowledges support from the Princeton IES. Code implementing the techniques introduced in this paper can be found at `https://github.com/rowanxshi/CDCP.jl`.

A combinatorial discrete choice problem (CDCP) is the problem of finding the combination of items that maximizes an objective function when the value of each item depends on which others are also chosen. The decision problem of a multinational firm selecting global production locations is an example: setting up production sites in Germany and the United States may generate scale economies that raise the value of both locations, while setting up such sites in Germany and France may lead to mutual cannibalization in the European market. Without imposing more structure on the objective function, the decision maker must evaluate all possible combinations, the number of which grows exponentially: with just twenty potential locations, there are already over one million combinations.

This paper develops a method to solve CDCPs whose objective function satisfies a single crossing condition—weaker than requiring items to be complements or substitutes—that discards suboptimal combinations without computing their payoffs. The method extends to heterogeneous-agent settings, always recovers the global optimum and, in our quantitative application, is orders of magnitude faster than evaluating every combination. Using it, we calibrate a general equilibrium model of multinational production in which firms choose foreign production locations. When locations are strongly complementary, gains from multinational production mainly accrue to the most productive firms and their headquarter countries, while some less productive countries even experience welfare losses; when they are substitutes, gains are positive and more even across countries.

Our method builds on Jia [2008], which shows that complementarity among choices enables an approach that discards suboptimal combinations without computing their payoffs. The key insight is that complementarity induces monotonicity: an item's value is highest when all others are chosen and lowest when it stands alone. If an item is valuable even in the least favorable environment—when chosen in isolation—it must always be included, and all combinations excluding it can be discarded; if it is not valuable even in the most favorable environment—when all other items are present—it must always be excluded, and all combinations including it can be discarded.

We show that this elimination logic extends to any CDCP satisfying a *single crossing differences* condition, either from below or from above. Single crossing differences from below generalizes complementarity: if an item's value is non-negative, it remains so as more items are added, though it may increase or decrease. Single crossing differences from above captures the symmetric case: if an item's value is non-negative, it remains so as items are removed, generalizing substitutability. For this latter case, the incumbent method is brute-force evaluation of all combinations, since the method of Jia [2008] does not apply.

Iteratively applying the elimination logic reduces a CDCP's domain by discarding combinations of items that cannot be optimal, but the reduced domain may still contain multiple candidate combinations. In such cases, Jia [2008] applies the brute-force approach to the reduced domain, evaluating all remaining combinations. We introduce a *branching* procedure that applies the elimination logic as far as possible before resorting to a brute-force step: when elimination can no longer make progress, it fixes one item's status, reapplies the elimination logic to each resulting case, and compares across cases to find the global optimum.

Heterogeneous agents compound the computational challenge: since each agent type may have a different optimal combination, solving the problem requires tackling a separate CDCP for every type in the population. For such settings, we extend our method to recover the exact *policy function* mapping each type to its optimal combination. This extension applies to CDCPs satisfying an additional single crossing condition: if including an item is worthwhile for a given agent type, it remains so for all higher types. When this condition holds, we can solve for the cutoffs at which the policy function changes value directly by using an indifference condition, avoiding the need to solve CDCPs for every agent type individually.

We apply our method to a general equilibrium model of multinational production in which each firm solves a CDCP when selecting production locations. Fixed costs of operating in each location prevent firms from producing everywhere, making the location choice combinatorial. Two counteracting channels determine whether locations are net complements or substitutes: an additional location increases firm-level scale economies but simultaneously cannibalizes demand across the firm's existing locations. We show that our single crossing differences conditions hold in the model, with locations being net complements when the scale parameter exceeds the cannibalization parameter, and net substitutes otherwise.

Our method makes it possible to calibrate the model to data on 32 OECD and European countries when locations are complements or substitutes: without it, each firm's location choice problem would require evaluating billions of combinations. The calibration targets bilateral trade flows, foreign affiliate sales, and affiliate counts across all country pairs, together with country-level moments on production, firm entry, and firm survival. Because production location sets vary across firms, the model does not yield aggregate gravity equations, so we use indirect inference, embedding our method inside the calibration loop to solve for policy functions at every parameter guess.

In benchmarking exercises using the calibrated model, our method outperforms existing approaches along three dimensions: computational speed, precision, and robustness across the range of complementarity and substitutability. We benchmark our method for up to 256 countries, sampling from our calibrated model to generate synthetic countries. Our method is up to four orders of magnitude faster than the brute-force approach and an order of magnitude faster than Jia [2008] with discretized firm heterogeneity when locations are complements, and retains its speed when locations are substitutes. On precision, we show that the incumbent approach of solving CDCPs for a discrete number of types and interpolating between them can induce substantial error in aggregate outcomes, while our approach is exact. Finally, our method is equally fast in both the complements and substitutes calibrations.

Whether locations are complements or substitutes determines whether welfare gains from multinational production concentrate in the most productive countries or accrue more broadly, a finding we establish by calibrating the model across more than 30 degrees of complementarity and substitutability. The stronger the complementarity, the more the distribution of welfare gains becomes uneven: countries home to large multinational firms, such as the United States, experience substantial gains, while some less productive countries that primarily host foreign production sites even lose. With substitutability, gains are positive and more even across countries since cross-location cannibalization limits the scale economies attainable by the most productive firms.

Modeling discrete choices when the value of an item depends on which others are chosen has long posed a challenge, compounded in settings with rich agent heterogeneity. The classic random utility framework of McFadden [1974] sidesteps it by restricting agents to choosing a single item, which rules out interactions among items but permits flexible heterogeneity through item-specific shocks and tractable aggregation. Extending the framework to allow agents to choose *multiple* items is tractable when items are independent [see, e.g., Hendel, 1999], but when they are interdependent, existing approaches restrict attention to settings with few items [see, e.g., Train et al., 1987, Gentzkow, 2007] due to the exponential growth in the number of combinations.

Our paper contributes to a growing literature on discrete choice problems with interdependent items outside the random utility framework. Jia [2008] introduces a solution method when choices are complements, which has been applied to store expansion, firm sourcing, and multinational production settings [see Antràs et al., 2017, Alfaro-Ureña et al., 2024, Antràs et al., 2024a].[1] We extend this line of work in two main ways. First, we provide a method that applies whenever items are substitutes rather than complements, and more generally whenever the objective function satisfies a single crossing condition. Second, we develop an exact heterogeneous-agent extension that avoids the approximation error induced by discretizing heterogeneity [see Tintelnot, 2017, Antràs et al., 2024a].

[1] A separate strand of this literature pursues solution methods that do not rely on monotonicity of choices. Computation-oriented approaches to CDCPs have used linear objective functions to apply integer programming [e.g., Head et al., 2026], greedy algorithms that do not guarantee global optimality [e.g., Fan and Yang, 2020], or machine learning methods to approximate heterogeneous-agent policy functions [e.g., Kulesza, 2024].

Relative to existing papers in the literature on multinational production, our tools make incorporating cross-location interactions and fixed costs possible, which we use to study how the welfare implications of multinational production depend on the presence of these forces. Existing quantitative work on multinationals has avoided CDCPs [see Ramondo and Rodríguez-Clare, 2013, Ramondo, 2014, Arkolakis et al., 2018], has solved small-scale CDCPs using the brute-force approach [see Tintelnot, 2017, Dyrda et al., 2024], or has restricted attention to cases with complementarities so that the method of Jia [2008] applies [see Antràs et al., 2017, Alfaro-Ureña et al., 2024, Antràs et al., 2024a]. Our counterfactual experiments suggest that interdependencies among production locations, such as scale economies and cannibalization effects, are central to understanding who gains and who loses from multinational production, with implications for the broader debate on the distributional consequences of globalization.

1. Defining and Solving Combinatorial Discrete Choice Problems

In this section, we formally define CDCPs and show how to solve them when the objective function satisfies two single crossing conditions. Let L denote a finite set of items indexed by ℓ, and $f : \mathcal{P}(L) \to \mathbb{R}$ the objective function, mapping the set of all possible combinations of items, $\mathcal{P}(L)$, to a real number. The Online Appendix contains a formal mathematical treatment of all statements and results.

1.1. Defining CDCPs

Consider a decision-maker who chooses a set $\mathcal{L} \in \mathcal{P}(L)$ to maximize f and denote the optimal decision set by $\mathcal{L}^\star \equiv \arg\max_{\mathcal{L}\in\mathcal{P}(L)} f(\mathcal{L})$. Our example throughout is a profit-maximizing firm selecting a set of foreign production locations $\mathcal{L}$ from $\mathcal{P}(L)$. Such decision problems are straightforward to solve when the value of any location $\ell \in L$ does not depend on the decision set $\mathcal{L}$: the firm can evaluate each location in isolation, reducing the problem to $|L|$ independent binary decisions. However, when the value of location ℓ depends on which other locations are in $\mathcal{L}$, the firm must evaluate combinations of locations jointly, making the problem *combinatorial*.

To formalize the notion of interdependence among locations, we introduce the following marginal value operator:

Definition 1 (Marginal Value Operator)**.** For all $\ell \in L$ and all $\mathcal{L} \in \mathcal{P}(L)$, the marginal value operator D_ℓ is

$$D_\ell f(\mathcal{L}) \equiv f(\mathcal{L} \cup \{\ell\}) - f(\mathcal{L} \setminus \{\ell\}).$$

We say that location choices are interdependent when the marginal value of at least a single $\ell \in L$ varies with the decision set $\mathcal{L}$. We now formally define CDCPs.

Definition 2 (Combinatorial Discrete Choice Problem)**.** A combinatorial discrete choice problem is the maximization problem

$$\max_{\mathcal{L} \in \mathcal{P}(L)} f(\mathcal{L}),$$

where there exist $\ell \in L$ and $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ such that $D_\ell f(\mathcal{L}) \neq D_\ell f(\mathcal{L}')$.

Without additional structure on the objective function f, no known algorithm exists to solve the generic CDCP in Definition 2 in polynomial time.[2] The brute-force approach evaluates f at all $2^{|L|}$ location combinations in $\mathcal{P}(L)$.

1.2. Solving CDCPs: A Single Crossing Differences Approach

In this section, we introduce a class of CDCPs whose objective function satisfies a single crossing condition and develop an iterative solution method that uses this structure to solve them without evaluating all combinations.

Single Crossing Differences in Choices Our solution method applies to CDCPs whose objective functions satisfy the following single crossing differences in choices (SCD-C) property.

Definition 3 (SCD-C)**.** The objective function f satisfies single crossing differences in choices from above if, for every $\ell \in L$ and for all decision sets $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ such that $\mathcal{L} \subset \mathcal{L}'$,

$$D_\ell f(\mathcal{L}') \geq 0 \quad \Rightarrow \quad D_\ell f(\mathcal{L}) \geq 0,$$

and single crossing differences in choices from below if

$$D_\ell f(\mathcal{L}) \geq 0 \quad \Rightarrow \quad D_\ell f(\mathcal{L}') \geq 0.$$

[2]The generic CDCP is NP-hard: no algorithm is known that solves all instances in time polynomial in $|L|$.

SCD-C restricts the marginal value of a location ℓ so that, along any sequence of expanding decision sets, it can cross zero at most once. SCD-C from below ensures that a location with a non-negative marginal value for some decision set $\mathcal{L}$ retains a non-negative marginal value as additional locations are added. SCD-C from above, conversely, ensures that a location with a non-negative marginal value for some $\mathcal{L}$ retains a non-negative marginal value as locations are removed.

Figure 1 illustrates the two SCD-C conditions. Both the solid and the dashed lines in the left panel show examples of how the marginal value of a location ℓ can vary as the location set expands from $\mathcal{L}_1$ to $\mathcal{L}_4$ when SCD-C from below holds. While the marginal value may change or even decrease, as with the solid line, it crosses zero at most once from below. The right panel shows two marginal value functions satisfying SCD-C from above, which can cross zero at most once from above.[3]

The properties of supermodularity and submodularity capture stronger notions of complementarity and substitutability and serve as sufficient conditions for SCD-C. A function f is supermodular if marginal values are increasing, that is, if $D_\ell f(\mathcal{L}) \leq D_\ell f(\mathcal{L}')$ for all $\mathcal{L} \subseteq \mathcal{L}'$ and all ℓ, which implies SCD-C from below. Submodularity reverses the inequality and implies SCD-C from above. While both the solid and dashed lines in Figure 1 depict marginal value functions consistent with SCD-C, only the dashed lines are consistent with super- or submodularity.

We illustrate the difference between SCD-C from below and the stronger supermodularity condition with an example. Consider a multinational firm that operates plants in Germany and the United States. Supermodularity requires that adding a plant in Canada raises the marginal value of both the German and US plants. In reality, however, the Canadian plant may lower the marginal value of the US plant—by cannibalizing its sales to the Canadian market—while raising the marginal value of the German plant by enhancing the firm's internal economies of scale. Such a scenario violates supermodularity, but it remains consistent with SCD-C from below as long as the US plant's marginal value stays weakly positive when the Canadian plant is added.

The Squeezing Procedure The squeezing procedure applies to CDCPs whose objective function satisfies either form of SCD-C. With SCD-C, a location's marginal

[3]Milgrom and Shannon [1994] introduce single crossing conditions into economics to derive comparative statics in settings without differentiability. There is a terminological discrepancy, acknowledged in Milgrom [2004], between Milgrom and Shannon [1994], which uses *single crossing condition*, and Milgrom [2004], which uses *single crossing differences*. The condition we refer to as SCD-C is similar to, but weaker than, the quasi-supermodularity condition of Milgrom and Shannon [1994]. We formally establish the relationship between these conditions in the Online Appendix. We adopt the single crossing *differences* terminology, to emphasize that the *marginal value* of the choice changes sign at most once.

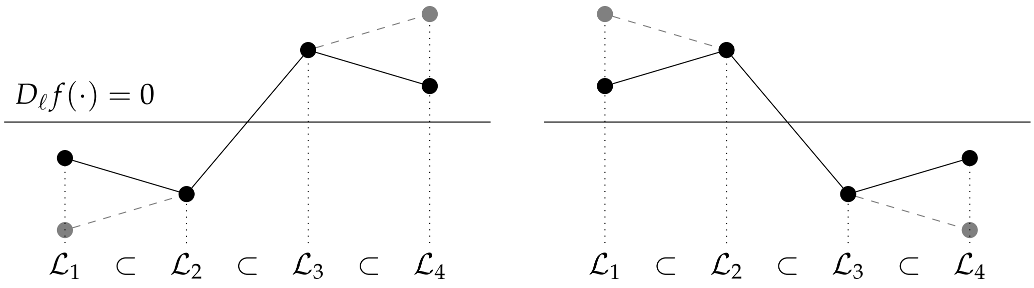


FIGURE 1: MARGINAL VALUE FUNCTIONS SATISFYING SCD-C

Note: Both panels show the marginal value of location ℓ evaluated at a sequence of expanding decision sets. In the left panel, the solid black line satisfies SCD-C from below but not supermodularity; the dashed line satisfies both. In the right panel, the solid black line satisfies SCD-C from above but not submodularity; the dashed line satisfies both. With SCD-C, marginal values can increase or decrease arbitrarily along the sequence as long as they cross zero at most once.

value crosses zero at most once, so that evaluating it at the extreme decision sets—the largest and smallest possible—can be sufficient to infer whether the location is optimal without evaluating any other decision sets. The procedure uses this logic to iteratively reduce the CDCP's domain, removing non-optimal decision sets from consideration without explicitly evaluating each one.

We introduce a pair of *bounding sets*, $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, comprising a lower and an upper bounding set satisfying $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}$. Each pair defines a subdomain of the original problem, and the original domain is represented by the *domain bounding sets*, $[\emptyset, L]$. We say that a pair of bounding sets $[\underline{\mathcal{K}}, \overline{\mathcal{K}}]$ is *tighter* than $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ if $\underline{\mathcal{L}} \subseteq \underline{\mathcal{K}}$ and $\overline{\mathcal{K}} \subseteq \overline{\mathcal{L}}$.

To solve a CDCP, we introduce a mapping that, starting with the domain bounding sets, produces progressively tighter bounding sets around the optimal decision set $\mathcal{L}^\star$, eliminating non-optimal sets from the domain. We refer to this mapping as the *squeezing step*.

Definition 4 (Squeezing step)**.** For all $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$, the squeezing step is the mapping defined by:

$$S\left([\mathcal{L}, \mathcal{L}']\right) \equiv \left[\inf\left\{\Phi(\mathcal{L}), \Phi(\mathcal{L}')\right\}, \sup\left\{\Phi(\mathcal{L}), \Phi(\mathcal{L}')\right\}\right],$$

where

$$\Phi(\mathcal{L}) \equiv \{\ell \in L \mid D_\ell f(\mathcal{L}) \geq 0\}.$$

At the heart of the squeezing step is the mapping Φ, which identifies the set of locations $\ell \in L$ with non-negative marginal value given the decision set $\mathcal{L}$. Importantly, $\Phi(\mathcal{L}^\star) = \mathcal{L}^\star$ by construction, so that the optimal decision set is a fixed point of the squeezing step.[4] Section 1.3 provides a detailed discussion of Jia [2008], which first introduced Φ to the economics literature.

When the objective function satisfies SCD-C, iteratively applying the squeezing step to the domain bounding sets generates a sequence of progressively tighter bounding sets that converges in at most $|L|$ steps. The following theorem formalizes this result.

Theorem 1. *If the objective function f exhibits SCD-C from above or below, iteratively applying the squeezing step to the domain bounding sets $[\emptyset, L]$ yields a sequence of bounding sets*

$$\emptyset = \underline{\mathcal{L}}^{(0)} \subseteq \cdots \subseteq \underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)} \subseteq \ldots \subseteq \overline{\mathcal{L}}^{(0)} = L,$$

where k denotes the iteration count. Iteration on the squeezing step converges in $K \leq |L|$ applications.

We now sketch a constructive proof of Theorem 1.[5] The mapping Φ is monotone if and only if the objective function f satisfies SCD-C, where the type of SCD-C determines the direction of the monotonicity. Consider two decision sets $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ such that $\mathcal{L} \subset \mathcal{L}'$. With SCD-C from above, if a location has a non-negative marginal value in the larger set $\mathcal{L}'$, it must also have a non-negative marginal value in any nested set $\mathcal{L}$. As a result, $\mathcal{L} \subset \mathcal{L}'$ implies $\Phi(\mathcal{L}') \subseteq \Phi(\mathcal{L})$: the mapping Φ is *order-reversing*. With SCD-C from below, the argument is similar: if a location has a non-negative marginal value in the smaller set $\mathcal{L}$, then it must have a non-negative marginal value in the larger set. As a result, $\mathcal{L} \subset \mathcal{L}'$ implies $\Phi(\mathcal{L}) \subseteq \Phi(\mathcal{L}')$ and the mapping Φ is *order-preserving*.[6]

When the type of SCD-C that f satisfies is known, the monotonicity of Φ simplifies the mapping S. With SCD-C from above, Φ is order-reversing: applying Φ to the upper bounding set yields a decision set nested in the one obtained from

[4] In the Online Appendix, we show that the set of fixed points of Φ contains at least one optimal decision set of f.

[5] The squeezing step tightens the bounding sets around the optimal decision set $\mathcal{L}^\star$ that includes locations for which the firm is indifferent. In the Online Appendix, we provide a way to check whether the identified optimum is unique, and a method to recover *all* the optima if it is not.

[6] We show the converse for the order-preserving case. Suppose Φ is order-preserving, and let $\mathcal{L} \subset \mathcal{L}'$ be arbitrary decision sets. Let ℓ be an arbitrary location. Now suppose $D_\ell f(\mathcal{L}) \geq 0$. Then, $\ell \in \Phi(\mathcal{L}) \subseteq \Phi(\mathcal{L}')$ since Φ is order-preserving; but $\ell \in \Phi(\mathcal{L}')$ implies $D_\ell f(\mathcal{L}') \geq 0$ by definition of Φ. As a result, Φ being order-preserving implies that $D_\ell f(\mathcal{L}) \geq 0 \Rightarrow D_\ell f(\mathcal{L}') \geq 0$. The proof in the order-reversing case is similar.

applying Φ to the lower bounding set. The squeezing step therefore simplifies to $S\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right) = \left[\Phi\left(\overline{\mathcal{L}}\right), \Phi\left(\underline{\mathcal{L}}\right)\right]$. With SCD-C from below, Φ is order-preserving, and the squeezing step becomes $S\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right) = \left[\Phi\left(\underline{\mathcal{L}}\right), \Phi\left(\overline{\mathcal{L}}\right)\right]$.

The monotonicity of Φ also ensures that the sets returned when applying the squeezing step to a pair of bounding sets are themselves always a valid pair of bounding sets. In particular, consider the bounding sets $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]$, so that $\underline{\mathcal{L}} \subseteq \mathcal{L}^{\star} \subseteq \overline{\mathcal{L}}$. If f satisfies SCD-C from above, applying Φ reverses this order, so $\Phi\left(\overline{\mathcal{L}}\right) \subseteq \Phi\left(\mathcal{L}^{\star}\right) \subseteq \Phi\left(\underline{\mathcal{L}}\right)$. Moreover, since $\Phi(\mathcal{L}^{\star}) = \mathcal{L}^{\star}$ by construction, this expression simplifies to $\Phi\left(\overline{\mathcal{L}}\right) \subseteq \mathcal{L}^{\star} \subseteq \Phi\left(\underline{\mathcal{L}}\right)$ so that $\left[\Phi\left(\overline{\mathcal{L}}\right), \Phi\left(\underline{\mathcal{L}}\right)\right]$ forms a new pair of bounding sets. Similarly, with SCD-C from below, the order-preserving nature of Φ implies $\Phi\left(\underline{\mathcal{L}}\right) \subseteq \mathcal{L}^{\star} \subseteq \Phi\left(\overline{\mathcal{L}}\right)$.

The monotonicity of Φ guarantees that, starting from the domain bounding sets $[\varnothing, L]$, iteratively applying the squeezing step produces a sequence of bounding sets that weakly tighten with each iteration. To see why, suppose f satisfies SCD-C from above, so that Φ is order-reversing. Applying the squeezing step to the domain bounding sets yields

$$\varnothing \subseteq \mathcal{L}^{\star} \subseteq L \qquad \Rightarrow \qquad \Phi\left(L\right) \subseteq \mathcal{L}^{\star} \subseteq \Phi\left(\varnothing\right).$$

The new pair of bounding sets $\left[\Phi\left(L\right), \Phi\left(\varnothing\right)\right]$ is vacuously tighter than the domain bounding sets $[\varnothing, L]$. Applying the squeezing step again produces:

$$\Phi\left(L\right) \subseteq \mathcal{L}^{\star} \subseteq \Phi\left(\varnothing\right) \qquad \Rightarrow \qquad \Phi\left(\Phi\left(\varnothing\right)\right) \subseteq \mathcal{L}^{\star} \subseteq \Phi\left(\Phi\left(L\right)\right).$$

The order-reversing property of Φ guarantees that the new upper bounding set $\Phi\left(\Phi\left(L\right)\right)$ is weakly tighter than the previous one, $\Phi\left(\varnothing\right)$: since $\varnothing \subseteq \Phi\left(L\right)$, it follows that $\Phi\left(\Phi\left(L\right)\right) \subseteq \Phi\left(\varnothing\right)$. By the same logic, the new lower bounding set is weakly tighter than the previous one. We conclude that

$$\varnothing \subseteq \Phi\left(L\right) \subseteq \Phi\left(\Phi\left(\varnothing\right)\right) \subseteq \mathcal{L}^{\star} \subseteq \Phi\left(\Phi\left(L\right)\right) \subseteq \Phi\left(\varnothing\right) \subseteq L.$$

Extending this logic inductively, each application of the squeezing step produces weakly tighter bounding sets. The procedure converges in at most $|L|$ steps, since each iteration either adds at least one location to the lower bounding set or removes at least one from the upper bounding set, and there are $|L|$ locations. An analogous argument holds when the objective function satisfies SCD-C from below. This completes our sketch of the proof of Theorem 1.

We refer to the iterative application of the squeezing step until convergence as the *squeezing procedure*, denoting the corresponding operator by S^K, and its output, $S^K\left([\varnothing, L]\right)$, as the *reduced* bounding sets. If the reduced bounding sets coincide at convergence, the optimal decision set has been found. If the reduced bounding sets do not coincide, they nevertheless define a weakly tighter subdomain of the original CDCP, on which we can identify $\mathcal{L}^\star$ either by brute-force evaluation of all remaining candidate decision sets or by a refinement procedure that we introduce next.

The Branching Procedure We introduce a recursive branching procedure to identify $\mathcal{L}^\star$ on a reduced domain $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]$ with $\underline{\mathcal{L}} \subset \overline{\mathcal{L}}$. The branching step selects a location ℓ from $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$, and creates two branches: the first imposes that ℓ is included in the firm's decision set, while the other imposes that ℓ is excluded. Each branch defines a *restricted* CDCP, specifically the problem of choosing $\mathcal{L} \subseteq \left(\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}\right) \setminus \{\ell\}$ to maximize the modified objective function $\tilde{f} : \mathcal{P}\left(\left(\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}\right) \setminus \{\ell\}\right) \to \mathbb{R}$ where:

$$\begin{aligned} \tilde{f}(\mathcal{L}) &\equiv f\left(\mathcal{L} \cup \underline{\mathcal{L}} \cup \{\ell\}\right) && \text{on the branch that includes } \ell, \\ \tilde{f}(\mathcal{L}) &\equiv f\left(\mathcal{L} \cup \underline{\mathcal{L}}\right) && \text{on the branch that excludes } \ell. \end{aligned}$$

The branching step applies the squeezing procedure until convergence to each restricted CDCP, yielding a pair of reduced bounding sets for each branch. The branching step then translates these bounding sets for the restricted CDCP into bounding sets for the original CDCP by adding all items in $\underline{\mathcal{L}}$ to both bounding sets on each branch. On the branch that includes ℓ, the item ℓ is also added to both bounding sets.

The output of the branching step is two pairs of conditional bounding sets for the original CDCP, one on each branch. If the bounding sets on a branch coincide, that branch has converged to a *conditionally* optimal decision set of the original CDCP; otherwise, the branching step is applied recursively. The recursive application of the branching step creates a tree in which each terminal node corresponds to a conditionally optimal decision set of the original CDCP.

We refer to the recursive application of the branching step until convergence on all branches as the *branching procedure* and define $\Lambda\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right)$ as the collection of conditionally optimal decision sets across the terminal nodes of all branches. The globally optimal decision set is $\mathcal{L}^\star = \arg\max_{\mathcal{L} \in \Lambda\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right)} f(\mathcal{L})$.[7]

Figure 2 illustrates the branching procedure: for a given pair of reduced bounding sets, two branches are created by applying the branching step with $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$. The branch on the left, which includes ℓ, converges to a pair of bounding sets with identical lower and upper bounds, identifying $\mathcal{L}_1^\star$ as the optimal decision set

[7] While $\Lambda\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right)$ may depend on the locations selected for branching, we show in the Online Appendix that it always contains the global maximum, so the branching procedure correctly identifies $\mathcal{L}^\star$ regardless of the branching order.

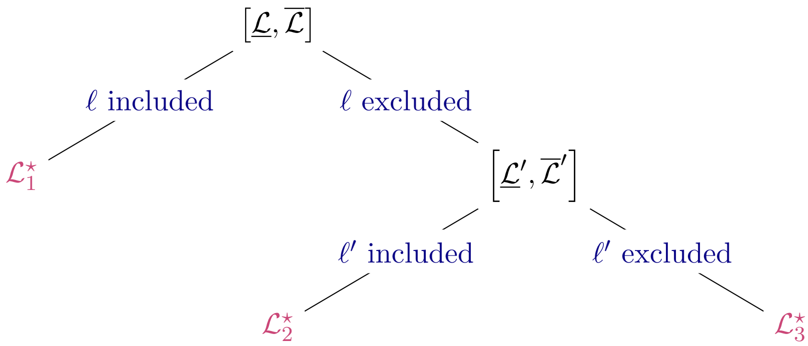


FIGURE 2: THE BRANCHING PROCEDURE

Note: This figure shows a tree of subproblems resulting from applying the branching procedure recursively. Convergence on a single branch occurs when the squeezing procedure returns a conditionally optimal set, indicated by a terminal node. The final output of the full recursive procedure is the collection of all conditionally optimal sets, in this example $\{\mathcal{L}_1^\star, \mathcal{L}_2^\star, \mathcal{L}_3^\star\}$.

conditional on including ℓ. The branch on the right, which excludes ℓ, converges to a pair of bounding sets which do not coincide, resulting in the bounding sets $\left[\overline{\mathcal{L}}', \underline{\mathcal{L}}'\right]$ on the original domain, so branching recurs: a new location $\ell' \in \overline{\mathcal{L}}' \setminus \underline{\mathcal{L}}'$ is selected, creating an additional pair of branches. This recursive process builds a tree, terminating in a conditionally optimal decision set on each branch so that $\Lambda\left([\underline{\mathcal{L}}, \overline{\mathcal{L}}]\right) = \{\mathcal{L}_1^\star, \mathcal{L}_2^\star, \mathcal{L}_3^\star\}$.

Branching is an alternative to the brute-force approach of evaluating the objective function at every decision set in the reduced domain. Intuitively, the branching procedure incorporates the brute-force logic one location at a time: it applies the squeezing procedure as much as possible, and when squeezing alone cannot identify the optimum, it fixes one location's status and repeats. In the worst case, when squeezing eliminates no decision set on any branch, the branching procedure evaluates the same decision sets as the brute-force approach.[8]

1.3. The Mathematics of Squeezing and Branching

This section explains the mathematical foundation of our squeezing method, formally connects it to Jia [2008], and identifies the new challenges that arise when the

[8]The wall-clock time advantage of branching over the brute-force approach depends on how effectively the squeezing procedure reduces the domain on each branch. When the squeezing procedure reduces the domain by little on each branch, the cost of repeated application may offset the savings from evaluating the objective function over fewer decision sets.

objective function satisfies SCD-C from above rather than below.

Jia [2008], Supermodularity, and SCD-C from Below Jia [2008] studies the location decisions of a retail chain, assuming individual store locations are complements so that the objective function is supermodular. In this context, Jia [2008] introduces the Φ mapping and recasts finding the optimal set of store locations, $\mathcal{L}^{\star}$, as finding the fixed points of Φ, since $\Phi(\mathcal{L}^{\star}) = \mathcal{L}^{\star}$ by construction.[9]

Since the objective function is supermodular, Φ is order-preserving and Jia [2008] invokes the theorem of Tarski [1955] to guarantee the existence of well-defined smallest and largest fixed points, $\mathcal{L}^{\inf}$ and $\mathcal{L}^{\sup}$. Together, these fixed points form a pair of bounding sets $\left[\mathcal{L}^{\inf}, \mathcal{L}^{\sup}\right]$, since necessarily $\mathcal{L}^{\inf} \subseteq \mathcal{L}^{\star} \subseteq \mathcal{L}^{\sup}$.

Jia [2008] identifies the bounding sets $\left[\mathcal{L}^{\inf}, \mathcal{L}^{\sup}\right]$ by applying Φ to $\emptyset$ and L separately until convergence. This approach is rooted in Kleene's fixed point theorem, which states that iteratively applying an order-preserving map to $\emptyset$ always converges to its smallest fixed point $\mathcal{L}^{\inf}$, while applying it to L always converges to its largest fixed point $\mathcal{L}^{\sup}$.[10] In cases where the bounding sets $\left[\mathcal{L}^{\inf}, \mathcal{L}^{\sup}\right]$ do not identify $\mathcal{L}^{\star}$, Jia [2008] applies the brute-force approach on the reduced domain.

In the case of SCD-C from below, our squeezing procedure implements the same method as Jia [2008]. With an order-preserving Φ, the squeezing step simplifies to $S\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right) = \left[\Phi\left(\underline{\mathcal{L}}\right), \Phi\left(\overline{\mathcal{L}}\right)\right]$, which is equivalent to applying Φ separately to the lower and upper bounding sets. Consequently, the reduced bounding sets produced by the squeezing procedure coincide with the smallest and largest fixed points of Φ, so that $S^{K}\left([\emptyset, L]\right) = \left[\mathcal{L}^{\inf}, \mathcal{L}^{\sup}\right]$.

Our first contribution relative to Jia [2008] is to extend the applicability of the paper's method beyond supermodular objective functions by showing that SCD-C from below is the necessary and sufficient condition for Φ to be order-preserving.

The Challenges of SCD-C from Above and Fixed Edges Our second contribution is to develop a solution method that applies when the objective function satisfies SCD-C from above. The key difficulty in this case is that Φ is order-reversing, so that the fixed point theorems of Tarski [1955] and Kleene do not apply; a smallest and largest fixed point bounding the optimal decision set may not exist.

[9]The mapping Φ has antecedents in the Operations Research literature on Boolean optimization. It presents a simple updating rule for decisions based on the discrete analogue of a first order condition [see, e.g., Boros and Hammer, 2002].

[10]See Stoltenberg-Hansen et al. [1994] for an exposition. The theorem derives from results in Kleene [1936] and Kleene [1938], though it was not stated explicitly there.

To see why, consider a firm choosing between two perfectly substitutable production locations, $L = \{\text{USA}, \text{CAN}\}$. Suppose that each location has positive marginal value when the other is inactive, but neither has positive marginal value when both are active, as each fully substitutes for the other. Applying the approach in Jia [2008] yields $\Phi(\emptyset) = \{\text{USA}, \text{CAN}\}$ and $\Phi(\{\text{USA}, \text{CAN}\}) = \emptyset$: iterating from either extreme oscillates and does not converge to a fixed point.

Although repeated application of Φ may not converge to a fixed point with SCD-C from above, it does converge to a pair of sets between which it alternates, as in the example above. We call a pair $[\mathcal{L}, \mathcal{L}']$ with $\mathcal{L} \subseteq \mathcal{L}'$ satisfying $\mathcal{L} = \Phi(\mathcal{L}')$ and $\mathcal{L}' = \Phi(\mathcal{L})$ a *fixed edge* of Φ. Fixed edges generalize the notion of fixed points: if $\mathcal{L}$ is a fixed point of Φ, then $[\mathcal{L}, \mathcal{L}]$ is a fixed edge.

Klimeš [1981] shows that, although an order-reversing map need not have smallest and largest fixed points, it always has a fixed edge $[\mathcal{L}^{\text{inf}}, \mathcal{L}^{\text{sup}}]$ that is extreme in the sense that $\mathcal{L}^{\text{inf}} \subseteq \mathcal{L} \subseteq \mathcal{L}' \subseteq \mathcal{L}^{\text{sup}}$ for all other fixed edges $[\mathcal{L}, \mathcal{L}']$.[11] The extreme fixed edge therefore serves as a pair of bounding sets, just as the smallest and largest fixed points do in the order-preserving case. With SCD-C from above, the squeezing procedure applied to the domain bounding sets converges to the extreme fixed edge of Φ.

The insights in this section imply that the squeezing step S is itself an order-preserving mapping whenever the underlying objective function satisfies SCD-C. An alternative proof of Theorem 1 therefore follows by applying the theorems of Tarski [1955] and Kleene to S directly.

The squeezing procedure converges to $[\mathcal{L}^\star, \mathcal{L}^\star]$ if and only if $\mathcal{L}^\star$ is the unique fixed point of Φ when f satisfies SCD-C from below, or the unique fixed edge of Φ when f satisfies SCD-C from above. When multiple fixed points or fixed edges exist, the branching procedure recovers all fixed points: every fixed point is a terminal node of the branching tree.

1.4. Solving CDCPs for Heterogeneous Agents

We now extend the framework to a setting with heterogeneous agents. We consider an objective function $f : \mathcal{P}(L) \times \mathbb{R} \to \mathbb{R}$ that maps a decision set $\mathcal{L} \in \mathcal{P}(L)$ and an agent type $z \in \mathbb{R}$ to a scalar payoff. We focus on single-dimensional heterogeneity

[11] The intuition behind Klimeš [1981] is simple. If Φ is order-reversing, then $\Phi^2 \equiv \Phi \circ \Phi$ is order-preserving. By Tarski [1955], Φ^2 has a smallest and largest fixed point. Any fixed point of Φ^2 satisfies $\Phi(\Phi(\mathcal{L})) = \mathcal{L}$, so the pair $[\mathcal{L}, \Phi(\mathcal{L})]$ is a fixed edge of Φ.

and assume that the objective function is continuous in z for all $\mathcal{L} \in \mathcal{P}(L)$.[12] Solving CDCPs for heterogeneous agents amounts to recovering the policy function that maps each agent's type to its optimal decision set.

Definition 5 (Policy function)**.** The policy function $\mathcal{L}^\star : \mathbb{R} \to \mathcal{P}(L)$ maps each type z to its optimal decision set, so that $\mathcal{L}^\star(z) \equiv \arg\max_{\mathcal{L} \in \mathcal{P}(L)} f(\mathcal{L}, z)$.

In our multinational production example, the policy function maps each firm's productivity to its optimal combination of production locations.[13]

We introduce a method to solve CDCPs with heterogeneous agents whenever the objective function satisfies SCD-C and an additional single crossing condition which we introduce next.

Single Crossing Differences in Type The single crossing differences in type (SCD-T) property restricts how the marginal value of a location varies with firm productivity.

Definition 6 (SCD-T)**.** The objective function f exhibits single crossing differences in type if, for all $\ell \in L$, decision sets $\mathcal{L} \in \mathcal{P}(L)$, and types $z, z' \in \mathbb{R}$ such that $z < z'$:

$$D_\ell f(\mathcal{L}, z) \geq 0 \qquad \Rightarrow \qquad D_\ell f(\mathcal{L}, z') \geq 0.$$

SCD-T ensures that if a location ℓ has a non-negative marginal value as part of a decision set $\mathcal{L}$ for a firm with productivity z, then it also has a non-negative marginal value for any firm with productivity $z' > z$. If instead the marginal value of ℓ crosses zero from above, the problem satisfies SCD-T with the transformed type $-z$.[14]

Just as SCD-C enables the squeezing procedure, which rules out decision sets without evaluating their payoff, SCD-T enables a *generalized* squeezing procedure that rules out decision sets for entire productivity ranges at once without evaluating their payoffs for any individual productivity value within those ranges.

[12] In an earlier working paper, we showed how to extend the methods in this section to settings with multidimensional heterogeneity [see Arkolakis et al., 2023].

[13] The policy function is well-defined when the optimal decision set is unique for each z except at cutoff productivity values. The Online Appendix provides a sufficient condition.

[14] This property is similar to, but weaker than, the single crossing property introduced by Milgrom and Shannon [1994], later referred to as the single crossing *differences* condition by Milgrom [2004]. We formally establish the relationship between these conditions in the Online Appendix. The firm type z can represent any characteristic, endogenous or exogenous, that affects payoffs, and the policy function $\mathcal{L}^\star(z)$ describes how optimal choices respond to changes in that characteristic. When z is endogenous, the policy function is often referred to as a best-response function.

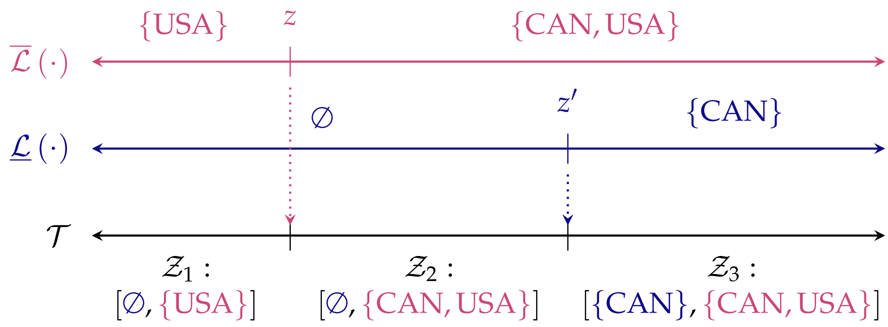


FIGURE 3: PARTITIONING THE PRODUCTIVITY SPACE BY COMMON BOUNDING SETS

Note: The middle line shows a lower bounding function, the top line an upper bounding function. Together, these two set-valued functions imply the partitioning $\mathcal{T}$, which creates intervals of productivity. In this figure, there are three intervals, so $\mathcal{T} = \{\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3\}$. All productivity values within an interval $\mathcal{Z}_t$ share the same bounding sets.

The Generalized Squeezing Procedure We extend the notion of bounding sets to *bounding functions*: set-valued functions of firm productivity $\underline{\mathcal{L}}(\cdot)$ and $\overline{\mathcal{L}}(\cdot)$, such that $\underline{\mathcal{L}}(z) \subseteq \mathcal{L}^{\star}(z) \subseteq \overline{\mathcal{L}}(z)$ for all $z \in \mathbb{R}$. We refer to $\left[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)\right] \equiv [\emptyset, L]$ as the domain bounding functions, since these constant functions nest the policy function for all z.

Any pair of bounding functions induces a partition of the firm productivity space into intervals over which the lower and upper bounding sets are constant. We define this partition as:

$$\mathcal{T}\left(\left[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)\right]\right) = \{\mathcal{Z}_1, \ldots, \mathcal{Z}_t, \ldots, \mathcal{Z}_T\}$$
$$\text{such that } \mathcal{Z}_t = \left\{z \in \mathbb{R} \mid \underline{\mathcal{L}}(z) = \underline{\mathcal{L}}_t, \overline{\mathcal{L}}(z) = \overline{\mathcal{L}}_t\right\},$$

where t indexes the intervals over which neither bounding function changes its value. When the bounding functions are clear from context, we write $\mathcal{T}$ for brevity.

Figure 3 illustrates how a pair of bounding functions partitions the productivity space into intervals. The middle and top lines represent the lower and upper bounding functions respectively, each changing value at different cutoffs, z and z'. The line at the bottom shows the resulting partition: the productivity intervals over which both bounding functions are constant. These intervals identify the cutoff productivity values at which either the lower or upper bounding function changes value.

We now define a generalized version of the squeezing step above.

Definition 7 (Generalized squeezing step)**.** The generalized squeezing step is the mapping S^g defined for all $\mathcal{L}(\cdot), \mathcal{L}'(\cdot) : \mathbb{R} \to \mathcal{P}(L)$ as:

$$S^g\left(\left[\mathcal{L}(\cdot), \mathcal{L}'(\cdot)\right]\right) \equiv \left[\inf\left\{\Phi^g\left(\mathcal{L}(\cdot), \cdot\right), \Phi^g\left(\mathcal{L}'(\cdot), \cdot\right)\right\}, \right. \\ \left. \sup\left\{\Phi^g\left(\mathcal{L}(\cdot), \cdot\right), \Phi^g\left(\mathcal{L}'(\cdot), \cdot\right)\right\}\right]$$

where the mapping $\Phi^g : \mathcal{P}(L) \times \mathbb{R} \to \mathcal{P}(L)$ is defined as

$$\Phi^g(\mathcal{L}, z) \equiv \left\{\ell \in L \mid z \geq z_\ell^g(\mathcal{L})\right\}$$

and the functions $z_\ell^g : \mathcal{P}(L) \to \mathbb{R}$ are defined as $z_\ell^g(\mathcal{L}) \equiv \inf\{z \in \mathbb{R} \mid D_\ell(\mathcal{L}, z) = 0\}$ for each $\ell \in L$.

The generalized squeezing step extends the logic of Definition 4 to entire ranges of productivity values at a time when the objective function satisfies SCD-C and SCD-T. For all $\mathcal{L}$ and z, $\Phi^g(\mathcal{L}, z)$ coincides with $\Phi(\mathcal{L})$ when the objective function is $f(\cdot, z)$. For all locations ℓ and decision sets $\mathcal{L}$, SCD-T implies that, rather than computing the marginal value of ℓ at each z, it suffices to check whether z lies above or below the cutoff $z_\ell^g(\mathcal{L})$ at which firms are indifferent to including ℓ in $\mathcal{L}$. Thus, the bounding functions update for entire productivity ranges at once, rather than point by point.[15]

When the objective function satisfies both SCD-C and SCD-T, iteratively applying the generalized squeezing step to the domain bounding functions returns a weakly tighter pair at each iteration. Iteration converges once $S^g\left(\left[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)\right]\right) = \left[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)\right]$. Since each bounding function can tighten at most $|L|$ times, iteration converges in at most $|L|$ applications. The following theorem formalizes this result.

Theorem 2. *If the objective function f exhibits SCD-C and SCD-T, iteratively applying the generalized squeezing step to the domain bounding functions $[\emptyset, L]$ yields a sequence of bounding functions so that for all $z \in \mathbb{R}$,*

$$\emptyset \subseteq \ldots \subseteq \underline{\mathcal{L}}^{(k)}(z) \subseteq \underline{\mathcal{L}}^{(k+1)}(z) \subseteq \mathcal{L}^\star(z) \subseteq \overline{\mathcal{L}}^{(k+1)}(z) \subseteq \overline{\mathcal{L}}^{(k)}(z) \subseteq \ldots \subseteq L,$$

where k denotes the iteration count. Iteration on the generalized squeezing step converges in $K \leq |L|$ applications.

We now illustrate how the generalized squeezing step proceeds for an objective function satisfying SCD-C from above and SCD-T. Consider the domain bounding functions, $[\emptyset, L]$, and the associated productivity partition $\mathcal{T}$. Since the domain

[15] If $D_\ell(\mathcal{L}, z) < 0$ for all z, we define $z_\ell^g(\mathcal{L}) \equiv \infty$; if $D_\ell(\mathcal{L}, z) > 0$ for all z, we define $z_\ell^g(\mathcal{L}) \equiv -\infty$. The range of each z_ℓ^g is thus the two-point compactification of the real line $\mathbb{R} \cup \{-\infty, \infty\}$.

bounding functions are constant for all z, the partition of the productivity space contains a single interval comprising the full productivity range. For each ℓ, we compute two cutoffs: $z_{\ell}^{g}(\emptyset)$, at which the marginal value of ℓ is zero in $\emptyset$, and $z_{\ell}^{g}(L)$ at which it is zero in L. The SCD-C and SCD-T conditions together imply $z_{\ell}^{g}(\emptyset) \leq z_{\ell}^{g}(L)$.

The indifference cutoffs $z_{\ell}^{g}(\emptyset)$ and $z_{\ell}^{g}(L)$ for each ℓ update the bounding functions as follows. For all firms with productivity $z < z_{\ell}^{g}(\emptyset)$, ℓ is not part of the optimal decision set, so the upper bounding function updates to $L \setminus \{\ell\}$. Conversely, for all firms with productivity $z > z_{\ell}^{g}(L)$, ℓ is part of the decision set, so the lower bounding function updates to $\{\ell\}$.

Figure 3 depicts the outcome of updating the domain bounding functions for $\ell =$ CAN. The figure shows the CAN-specific cutoffs $z_{\text{CAN}}^{g}(\emptyset)$ as z and $z_{\text{CAN}}^{g}(L)$ as z'. The updated bounding functions now vary with productivity, changing value at $z_{\text{CAN}}^{g}(\emptyset)$ and $z_{\text{CAN}}^{g}(L)$, and hence partition the productivity space into three intervals. A full update of the bounding functions requires computing such cutoffs for all locations ℓ. In subsequent applications, the generalized squeezing step applies to each partition interval separately, returning a finer partition and thus tightening the bounding functions around the policy function.

We refer to the iterative application of the generalized squeezing step until convergence as the *generalized squeezing procedure* and the resulting tightened bounding functions as *reduced* bounding functions. The reduced bounding functions pin down the policy function on any interval for which they coincide.

For productivity intervals for which bounding functions differ, Appendix A defines two refinement methods that converge to the policy function on all intervals of the productivity space. The first, the *generalized branching procedure*, extends the branching procedure from Section 1.2 to the heterogeneous-agent case. The second, *iterative cutoff search*, requires the objective function to satisfy a stronger version of SCD-T, which we call *strong SCD-T*. Strong SCD-T requires that, for any pair of decision sets $\mathcal{L}$ and $\mathcal{L}'$, the difference in payoffs $f(\mathcal{L}, z) - f(\mathcal{L}', z)$ crosses zero for at most one value of z. Strong SCD-T implies SCD-T, but not vice versa: SCD-T only requires such single crossing for pairs $(\mathcal{L} \setminus \{\ell\}, \mathcal{L} \cup \{\ell\})$ that differ by a single element.[16] Strong SCD-T, which holds in the application of Section 2, imparts additional structure on the policy function that iterative cutoff search leverages.

We refer to the application of generalized squeezing and its refinement until convergence as the *policy function method*. In contrast to discretization or bisection methods, the policy function method uses the monotonicity afforded by SCD-T to

[16]SCD-T also allows the crossing at zero to occur for an interval of productivity values instead of at a single value; strong SCD-T rules out this possibility.

identify the exact cutoffs at which the policy function changes value. This approach eliminates unnecessary computation at non-cutoff productivity values and avoids the interpolation required by discretization methods. Section 4 shows that such interpolation can introduce substantial errors in aggregate variables in the context of a calibrated multinational production model.

2. A Quantitative Model of Multinational Production

In this section, we introduce a quantitative model of multinational production (MP) and derive conditions under which our solution method applies. In the model, a firm's problem of choosing multinational production locations maps directly into the CDCP framework of Section 1: the set of candidate locations L corresponds to the set of countries, the objective function f to the firm's operating profits, and the agent type z to firm productivity. The single crossing conditions translate into restrictions on the demand and cost elasticities. In the Online Appendix, we extend the framework to accommodate a broader class of demand and cost functions.

Environment The world economy consists of a discrete set of countries L. Firms are headquartered in one country, produce in a set of countries, and sell to consumers in all countries. We index headquarter locations by i, production locations by ℓ, and final consumption destinations by n. Each firm produces a differentiated final good variety with a firm-specific productivity z. Since firms with the same headquarter location and productivity behave identically, we index them by i and z alone. In every location ℓ, a mass of H_ℓ households inelastically supplies one unit of labor at wage w_ℓ. Labor markets are perfectly competitive and output markets are monopolistically competitive.

Demand System Consumers in all destinations have identical CES preferences with elasticity $\sigma > 1$ over the set of available final goods. The demand function for an individual variety as a function of its destination-specific price p_n is

$$q_n\left(p_n\right) = Q_n \left(\frac{p_n}{P_n}\right)^{-\sigma},$$

where Q_n is the CES consumption basket and P_n the associated ideal price index. Firms and consumers take the aggregate objects Q_n and P_n as given.

Production Technology Consider a firm headquartered in location i with productivity z that operates a set of production locations $\mathcal{L} \subseteq L$. The firm delivers its final good to destination n by combining production from its locations ℓ. The firm's marginal cost of serving destination n is given by a CES aggregator of the marginal cost at each of its production locations:

$$(1) \qquad c_{in}(\mathcal{L}, z) = \frac{1}{z}\left[\sum_{\ell \in \mathcal{L}} \xi_{i\ell n}^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}, \qquad \xi_{i\ell n} \equiv \gamma_{i\ell}\frac{w_\ell}{A_\ell}\tau_{\ell n}.$$

The parameter $\varepsilon > 1$ denotes the elasticity of substitution among production locations. For each production location, w_ℓ is the equilibrium wage rate and A_ℓ is an exogenous location-specific productivity shifter. Firms also face a bilateral MP cost $\gamma_{i\ell}$, which captures factors such as communication or language costs, as well as a bilateral cost of trade $\tau_{\ell n}$. We summarize all cost shifters in a trilateral resistance term denoted $\xi_{i\ell n}$.

Equation (1) shows that, all else equal, expanding the production location set lowers a firm's marginal cost. However, the cost savings from each additional location fall as the production location set grows. As $\varepsilon \to \infty$, firms only use their lowest-cost location.

The marginal cost aggregator in equation (1) captures, in reduced form, that firms use output from all their production locations to assemble their final good for any destination. One microfoundation is that, for each destination market, a firm's locations compete to be the lowest-cost supplier for each of a continuum of inputs into the final product. Adding a production location expands the set of potential suppliers and lowers the firm's marginal cost, but at a diminishing rate: as the location set grows, the probability that any given location is the lowest-cost supplier for a specific input declines. In the Online Appendix, we detail this microfoundation with Fréchet-distributed cost shocks, as in Antràs et al. [2017] and Tintelnot [2017].[17] The equilibrium allocations in our model depend only on the aggregator in equation (1) and not on the underlying microfoundation.

Profit Maximization Consider a firm headquartered in location i with productivity z. The firm's profit maximization problem has two parts: choosing a set of production locations $\mathcal{L}$ and setting the price of its final good in each destination

[17] Antràs et al. [2017] model firms that pay fixed costs to add sourcing partners subject to Fréchet cost shocks, while Tintelnot [2017] studies firms that establish plants across locations with plant-specific Fréchet shocks. Tintelnot [2017] restricts the Fréchet shape parameter to exceed a common demand and cost elasticity, which implies SCD-C from above. Our framework allows the two elasticities to differ. Alternative foundations include multivariate Pareto shocks [Arkolakis et al., 2018] or intermediate goods that are imperfect substitutes across locations [Antràs et al., 2024b].

market. For a given $\mathcal{L}$, the firm sets its price in destination n as a constant markup over marginal cost, $c_{in}(\mathcal{L}, z)$, so that

$$p_{in}(\mathcal{L}, z) = \frac{\sigma}{\sigma - 1} c_{in}(\mathcal{L}, z).$$

Given its optimal pricing policy, the firm chooses the optimal location decision set $\mathcal{L}_i^\star(z)$ to maximize its operating profits:

$$\text{(2)} \qquad \max_{\mathcal{L} \in \mathcal{P}(L)} \pi_i(\mathcal{L}, z) \equiv \max_{\mathcal{L} \in \mathcal{P}(L)} \left\{ \sum_{n \in L} \frac{1}{\sigma} q_n(p_{in}(\mathcal{L}, z)) \, p_{in}(\mathcal{L}, z) - \sum_{\ell \in \mathcal{L}} w_\ell f_{i\ell} \right\},$$

where $f_{i\ell}$ is the fixed cost of producing in location ℓ, denominated in location ℓ production labor. Note that, with CES preferences, variable profits are a share $1/\sigma$ of revenue.

Finally, the firm must first pay a labor-denominated entry cost f_i^e to draw a productivity z from an exogenous distribution $G_i(z)$. The firm produces positive quantities if it has non-negative operating profits; otherwise, it exits right after paying f_i^e.

Aggregation and the Equilibrium System We now turn to aggregation and the equilibrium determination of aggregate variables.

The first equilibrium condition is a zero profit condition that pins down the cutoff productivity $\tilde{z}_i$ below which firms would have negative operating profits and thus exit instead:

$$\text{(3)} \qquad \pi_i(\mathcal{L}_i^\star(\tilde{z}_i), \tilde{z}_i) = 0.$$

The second equilibrium condition is a free entry condition that pins down the total mass of entrants M_i in each origin country i by equalizing the entry cost on the left with the expected operating profits on the right:

$$\text{(4)} \qquad w_i f_i^e = \int_{\tilde{z}_i}^{\infty} \pi_i(\mathcal{L}_i^\star(z), z) \, \mathrm{d}G_i(z).$$

Price indices in each destination market n aggregate the prices of all goods offered:

$$\text{(5)} \qquad P_n^{1-\sigma} = \sum_{i \in L} M_i \int_{\tilde{z}_i}^{\infty} p_{in}(\mathcal{L}_i^\star(z), z)^{1-\sigma} \, \mathrm{d}G_i(z).$$

In addition, the labor market must clear in each production location ℓ. There are three sources of labor demand: variable labor requirements from all the production sites operating in country ℓ, the entry costs of all the firms headquartered in ℓ, and the fixed costs incurred by foreign and domestic firms to set up production in location ℓ. Labor market clearing equates the total labor supply in location ℓ to total labor demand as follows:

$$w_\ell H_\ell = \frac{\sigma-1}{\sigma} \sum_{i\in L} \sum_{n\in L} M_i X_n \int_{\tilde{z}_i}^{\infty} \frac{\mathbb{1}_{i\ell}(z)\, \xi_{i\ell n}^{1-\varepsilon}}{\sum_{k\in\mathcal{L}_i^\star(z)} \xi_{ikn}^{1-\varepsilon}} \left[\frac{p_{in}\left(\mathcal{L}_i^\star(z), z\right)}{P_n}\right]^{1-\sigma} \mathrm{d}G_i(z) \tag{6}$$

$$+ M_\ell w_\ell f_\ell^e + \sum_{i\in L} M_i w_\ell f_{i\ell} \int_{\tilde{z}_i}^{\infty} \mathbb{1}_{i\ell}(z)\, \mathrm{d}G_i(z),$$

where X_n denotes total expenditure on final goods in destination n and the indicator $\mathbb{1}_{i\ell}(z) \equiv \mathbb{1}\{\ell \in \mathcal{L}_i^\star(z)\}$ is 1 if a firm headquartered in i with productivity z opens a production location in country ℓ.[18]

Lastly, balance of payments requires that total expenditure from consumers in a market n equals their total income:

$$X_n = w_n H_n. \tag{7}$$

The general equilibrium in our model is defined as follows.

Definition. General equilibrium is a set of policy functions $\{\mathcal{L}_i(\cdot)\}_i$ and aggregate variables $\{w_i, \tilde{z}_i, M_i, P_i, X_i\}_i$ so that

1. given the aggregate variables, the policy functions solve the firm optimization problem in equation (2); and
2. given the policy functions, the aggregate variables satisfy equations (3)–(7).

Establishing the SCD-C and SCD-T Conditions We conclude this section by establishing the SCD-C and SCD-T properties in the model.

To build intuition, consider a symmetric version of our economy where $\xi_{i\ell n} = \xi$, so that the marginal value of an additional production location depends only on the number of active locations, not on their identity:

$$D_\ell \pi_i(\mathcal{L}, z) = \mathcal{G} z^{\sigma-1} \left(|\mathcal{L} \cup \{\ell\}|^{\frac{\sigma-1}{\varepsilon-1}} - |\mathcal{L} \setminus \{\ell\}|^{\frac{\sigma-1}{\varepsilon-1}} \right) - w_\ell f_{i\ell}, \tag{8}$$

where $\mathcal{G}$ is a composite general equilibrium constant. A single sufficient statistic, the ratio $(\sigma-1)/(\varepsilon-1)$, governs whether locations are complements or substitutes

[18] Note that, with CES demand, a fraction $(\sigma-1)/\sigma$ of a firm's total sales are variable production costs; location $\ell \in \mathcal{L}_i^\star(z)$ accounts for a share $\mathbb{1}_{i\ell}(z)\, \xi_{i\ell n}^{1-\varepsilon} / \sum_{k\in\mathcal{L}_i^\star(z)} \xi_{ikn}^{1-\varepsilon}$ of these costs.

and determines the type of SCD-C the profit function satisfies. When $(\sigma - 1)/(\varepsilon - 1) > 1$, the marginal value of any location is increasing in the number of other locations, so that supermodularity and hence SCD-C from below holds. When $(\sigma - 1)/(\varepsilon - 1) < 1$, the marginal value of any location is decreasing in the number of other locations, so that submodularity and hence SCD-C from above holds. When the ratio is one, the marginal value is independent of the firm's decision set $\mathcal{L}$.[19]

The ratio $(\sigma - 1)/(\varepsilon - 1)$ summarizes the net effect of two opposing forces that generate interdependencies among production locations. On the demand side, the elasticity of substitution $\sigma > 1$ governs the firm's ability to scale: since demand is price-sensitive, the same price reduction translates to a larger sales volume if the firm is already low-price, so that each additional production location raises the value of the next. On the cost side, since $\varepsilon > 1$, production locations are substitutes, as they effectively compete with one another to supply each destination. When $(\sigma - 1)/(\varepsilon - 1) > 1$, the demand side complementarity dominates the supply side substitutability.

Equation (8) also shows that SCD-T is always satisfied in our model since the marginal value of location ℓ is a strictly increasing function of the firm's productivity type given $\sigma > 1$. Moreover, this parameter restriction ensures strong SCD-T: for any two decision sets, there is at most one productivity at which the firm is indifferent between them.

As we show formally in the Online Appendix, these parameter conditions remain the sufficient conditions for SCD-C and strong SCD-T in the full model, so that the policy function method with iterative cutoff search is applicable.[20]

[19]Equation (8) highlights that the firm's CDCP arises from the combination of two ingredients: interdependencies among locations and the fixed costs of setting up production locations. Without interdependencies, that is, when $(\sigma - 1)/(\varepsilon - 1) = 1$, the marginal value of any production location is independent of all others and hence the firm faces $|L|$ independent decisions. Without fixed costs, that is, when $f_{i\ell} = 0 \, \forall i, \ell$, the marginal value of all locations is always non-negative, so firms establish production locations in all countries, regardless of interdependencies.

[20]In the Online Appendix, we also show how to verify SCD-C and SCD-T for more general demand functions $q_n(p_n)$ and marginal cost functions $c_{in}(\mathcal{L}, z)$ of which the CES functions in this section are special cases. Our generalized framework nests demand systems as in Arkolakis et al. [2019] and production structures as in Ramondo [2014], Arkolakis et al. [2018], and Lind and Ramondo [2023]. In general, the strength of the complementarities or substitutabilities across locations depends on how changes in marginal costs translate into differences in variable profits, which is determined by the elasticity of demand and the pass-through elasticity of costs to price. In the present model, markups are constant so that the pass-through elasticity is always 1; consequently, only the elasticity of demand matters.

3. Quantification

We calibrate the model to data on trade and MP across 32 countries. As part of the calibration routine, we solve the full heterogeneous-agent general equilibrium model for each parameter guess, a task our policy function method makes computationally feasible. This section provides an overview of the calibration strategy, with further details on how it uses our solution methods in Appendix C. The Online Appendix discusses data treatment and model fit.

3.1. Data

We use information on manufacturing trade and MP for 32 OECD and European countries from Alviarez [2019]. For each host country, the dataset contains the number of foreign affiliates in manufacturing and their total sales by origin country. For example, the dataset contains the number of Canadian affiliates of German firms and their total sales. The dataset also contains bilateral trade flows for all country pairs. For all variables, the data reflect average values over the period from 2003 to 2012.[21]

While the dataset contains foreign affiliate counts for each country, it lacks total firm counts and information on firm entry and survival. To supplement the dataset, we obtain total firm counts for each country from the UNIDO Industrial Statistics Database, and data on the one-year survival rates of firms in each country from the OECD Structural and Demographic Business Statistics.

To construct standard bilateral gravity controls, we use the CEPII database [see Conte et al., 2023], which provides measures of geographic distance, a common border indicator, a shared colonial past indicator, and a shared language indicator. In addition, we obtain tariff data from the Global Tariff database [see Teti, 2024] and data on real GDP per capita and total employment from the Penn World Table [see Feenstra et al., 2015].

3.2. Calibration Strategy

Table 1 summarizes our calibration strategy, listing each parameter and the moment it targets. We set σ and ε externally and then use country-level aggregate outcomes together with bilateral trade, MP, and affiliate count flows to discipline the remaining parameters.

[21] To construct the dataset, Alviarez [2019] combined data from the OECD, the Eurostat Foreign Affiliate Statistics database, the Bureau of Economic Analysis, and Bureau van Dijk's Orbis dataset.

TABLE 1: PARAMETERS AND TARGET MOMENTS

PARAMETER	DESCRIPTION	VALUE/TARGET MOMENT
Set Externally		
σ	Demand elasticity	Set to 4 [Arkolakis et al., 2018]
$\frac{\sigma-1}{\varepsilon-1}$	Interdependencies among locations	Set from 0.25 to 3.9 with benchmark values $\frac{3}{2}$ and $\frac{2}{3}$
Calibrated Internally		
ζ	Firm Pareto shape	Sales right-tail [Arkolakis, 2010]
$\underline{z}_i$	Firm Pareto minimum	Share of global MP conducted by firms HQed in i [Alviarez, 2019]
A_ℓ	Location productivity	GDP per capita (PWT)
f_i	Fixed cost (origin)	One-year firm survival rate (OECD)
f_i^e	Entry cost	Total firm counts (UNIDO)
H_ℓ	Labor supply	Employment (PWT)
$\tau_{\ell n}, \gamma_{i\ell}, \nu_{i\ell}$	Bilateral trade, MP, fixed costs	Gravity coefs. on trade flows, MP flows, affiliate counts [Alviarez, 2019]

Note: This table summarizes the calibration strategy of the model. Two parameters, σ and ε, are set externally; the remaining are calibrated simultaneously in a single iterative routine that targets the moments described in the last column, subject to general equilibrium conditions. For each empirical moment, we list the data source. PWT refers to the Penn World Table, OECD to the OECD Structural and Demographic Business Statistics, and UNIDO to the UNIDO Industrial Statistics Database.

Demand Elasticity and Location Substitution Elasticity: σ and ε The ratio $(\sigma - 1)/(\varepsilon - 1)$ determines whether locations are complements or substitutes, as well as the strength of these forces. We set $\sigma = 4$, following Arkolakis et al. [2018].[22] To cover a wide range of complementarity and substitutability, we calibrate the model for 37 values of $(\sigma - 1)/(\varepsilon - 1)$, evenly spaced between 0.25 and 3.9, which implicitly defines a range of values for ε. We choose two values of this ratio as the benchmarks for our quantitative exercises below, one for the complements case and one for the substitutes case. The substitutes benchmark is $(\sigma - 1)/(\varepsilon - 1) = 2/3$, corresponding to $\varepsilon = 5.5$ as in Arkolakis et al. [2018], while the complements benchmark is $(\sigma - 1)/(\varepsilon - 1) = 3/2$, chosen for symmetry.

[22]Our choice of σ falls within the range of estimates of Broda and Weinstein [2006]. Arkolakis et al. [2018] show that it is also consistent with markup estimates from the manufacturing sector. Moreover, our value is similar to the estimate of $\sigma = 3.89$ from Head and Mayer [2019]. While Head and Mayer [2019] focuses on MP in the car industry, their estimation strategy is consistent with our theoretical setup and also generates markups in line with the microeconomic estimates from the car industry in Goldberg [1995] and Berry et al. [1995].

Country Productivity Parameters and Fixed Costs We assume that the firm productivity distribution $G_i(z)$ is Pareto with shape parameter ζ and minimum $\underline{z}_i$, which is specific to each headquarter country i. The model generates a firm sales distribution with a Pareto tail of $\zeta/(\sigma-1)$, and we set $\zeta = 1.65 \times (\sigma - 1)$ to match the tail estimate in Arkolakis [2010].

The model features two vectors of country-specific productivity terms: the minimum of the firm Pareto distribution, $\underline{z}_i$, which acts as a headquarter-location productivity shifter; and the production-location productivity shifter, A_ℓ. We choose the first to match the share of global foreign production attributable to firms headquartered in i, and the second to match the observed GDP per capita for each country.

For the calibration, we decompose the fixed cost of setting up a production location in ℓ for firms headquartered in location i into an origin-specific shifter and a bilateral term, $f_{i\ell} \equiv f_i \nu_{i\ell}$. We choose the base component of fixed costs f_i to match the observed one-year firm survival rate and the entry cost f_i^e to match the mass of entering firms. In the data, we compute the total mass of entrants as a country's total firm count divided by its one-year firm survival rate. We set total labor supply in each location, H_ℓ, to match total employment in the country.

Trade Costs, MP Costs, Fixed Costs In the theory, three bilateral cost matrices shape the patterns of trade flows, foreign affiliate sales, and foreign affiliate counts across country pairs: the matrix of trade costs $\{\tau_{\ell n}\}_{\ell n}$, the matrix of MP costs $\{\gamma_{i\ell}\}_{i\ell}$, and the matrix of the bilateral components of fixed costs $\{\nu_{i\ell}\}_{i\ell}$. Following Tintelnot [2017], we parameterize these matrices as constant elasticity functions of the standard bilateral gravity variables: geographic distance, a shared colonial past indicator, a shared language indicator, and a common border indicator. Each matrix has a separate elasticity for every gravity variable, totaling twelve elasticities. We additionally include bilateral tariffs in the trade cost matrix, with an elasticity of one. We present the expressions for all three bilateral cost matrices in the Online Appendix.

We identify the elasticities using an indirect inference approach, since the model does not deliver closed-form aggregate gravity equations. Using trade flows, MP flows, and bilateral affiliate counts from the model, we estimate separate gravity equations for each and choose the elasticities to match the corresponding regression coefficients in the data.[23]

[23] For any bilateral pair without MP activity in the data, we set $\gamma_{i\ell} = \infty$.

4. Quantitative Exercises and Counterfactuals

This section applies our solution method to the calibrated model. We first demonstrate the method's computational advantages over incumbent approaches, then examine how the distribution of welfare effects from MP depends on whether production locations act as complements or substitutes. For each exercise, we use either our benchmark complements and substitutes calibrations, or the full range of calibrated economies differing in the degree to which locations are complements or substitutes.

4.1. Computational Performance

We conduct three numerical experiments, demonstrating the advantages of our solution method in terms of speed, precision, and breadth of applicability. For experiments that vary the number of countries, we sample from the location-specific parameters of our calibrated model to generate synthetic countries.[24]

Speed We compare our policy function method ("Policy"), which combines generalized squeezing with the iterative cutoff search refinement, against two alternative approaches that discretize firm productivity on a grid of $2^{14} \approx 16000$ points. The "Naive" approach solves the CDCP at each grid point using the brute-force approach of computing the profits for all possible location combinations. The "Squeezing" approach applies the squeezing procedure and then applies the brute-force approach to the reduced domain. Both alternatives use a productivity grid and therefore produce interpolation error in aggregation. Our policy function method is exact and hence avoids a grid entirely.

Table 2 reports the time (in seconds) required to compute the policy function across varying numbers of synthetic countries. We compute $\mathcal{L}_i^\star(\cdot)$ for firms from each origin i and report the average time across origins. With just 16 countries, the naive method requires more than an hour, while the policy method solves the same problem in less than a tenth of a second. Even with 128 countries, the policy method recovers the average policy function in under 15 seconds, and with 256 countries in under 10 minutes.

[24] A synthetic country is described by three types of parameters: fundamentals $\{A_\ell, f_i, H_\ell\}$, aggregates $\{w_\ell, P_\ell\}$, and bilateral costs $\{\tau_{\ell n}, \gamma_{i\ell}, \nu_{i\ell}\}$. For each, we non-parametrically fit a distribution to the estimated values and sample from it. We draw 256 synthetic countries this way, then order them by their market size, $w_\ell H_\ell$. In Table 2 and Figure 4, each numerical exercise subsets from this collection of synthetic countries, from the largest to smallest.

TABLE 2: RUNTIMES (SECONDS) FOR DIFFERENT SOLUTION METHODS

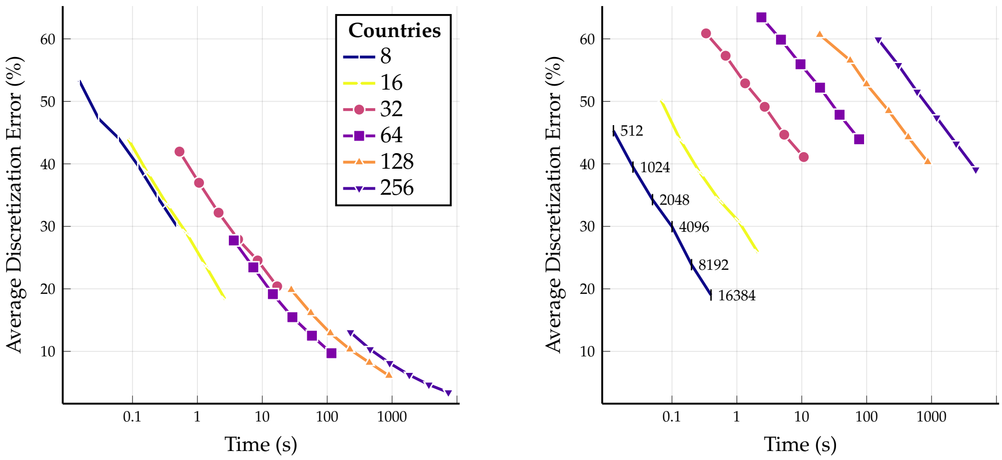
Countries	Complements			Substitutes		
	Naive (1)	Squeezing (2)	Policy (3)	Naive (1)	Squeezing (2)	Policy (3)
8	6	0.482	0.034	11	0.400	0.009
16	5215	2.699	0.087	8444	2.142	0.018
32	–	16.968	0.186	–	10.735	0.118
64	–	116	1.293	–	76.540	1.324
128	–	895	14.702	–	874	14.167
256	–	7375	374	–	4797	565
Grid points	2^{14}	2^{14}	–	2^{14}	2^{14}	–

Note: This table reports the time (in seconds) to compute the firm policy function $\mathcal{L}_i^\star(\cdot)$ using three methods. The "Naive" method discretizes firm productivity on a grid and evaluates the profit function for all possible production location combinations at each grid point. The "Squeezing" method applies the squeezing procedure at each grid point until convergence, then uses the brute-force approach on the reduced domain. Our "Policy" method solves for the exact policy function without discretization. Results are shown for both the complements and substitutes calibrations. Synthetic countries are generated by sampling from the distribution of parameter estimates obtained from the calibrated model. We report the average time to solve for $\mathcal{L}_i^\star(\cdot)$ across all origins i. All computations were performed on an Apple M1 (2020) CPU.

In the complements case, the squeezing method corresponds to the incumbent approach pioneered by Antràs et al. [2017], and our policy function method is one to two orders of magnitude faster across all country counts. In the substitutes case, the naive brute-force approach represents the incumbent approach, and our method improves on it by four to five orders of magnitude. Approximately half of the speed gains come from extending the squeezing method of Jia [2008] to the substitutes case, and the remaining half from introducing our policy function method, which avoids solving the CDCP at every grid point.

Table 2 does not account for the discretization error of the alternative approaches; we quantify the size of this error next.

Precision Our policy function method solves for the *exact* policy function, allowing us to quantify the discretization error of other methods, which arises because decision sets must be interpolated between grid points.



(A) Complements

(B) Substitutes

FIGURE 4: THE PRECISION-TIME FRONTIER WITH DISCRETIZATION METHODS

Note: This figure shows the trade-off between computation time and discretization error for the "Squeezing" method, for varying numbers of countries. Discretization error is measured as the average percentage deviation of trilateral flows computed using the discretized policy function from those computed using the exact policy function obtained by the "Policy" method. Specifically, we compute $N^{-3} \sum_{i,\ell,n \in L} \left| \tilde{X}_{i\ell n} / X_{i\ell n} - 1 \right| \times 100\%$, where $X_{i\ell n}$ denotes sales to destination n by production sites in ℓ whose headquarters are in i, computed using the exact policy function, and $\tilde{X}_{i\ell n}$ is the corresponding object computed using the "Squeezing" method with interpolation between grid points. All computations were performed on an Apple M1 (2020) CPU.

To quantify the trade-off between computational time and discretization error inherent in grid-based approaches, we first compute the total value of trade flows from origin i to destination n via production in country ℓ, denoted $X_{i\ell n}$, by aggregating over the exact policy function. We then compute the corresponding approximation $\tilde{X}_{i\ell n}$ by aggregating over the discretized policy function of the "Squeezing" approach, interpolating between grid points. We measure discretization error as the average absolute percentage deviation between flows computed from the discretized policy function and the exact policy function, that is, $N^{-3} \sum_{i,\ell,n \in L} \left| \tilde{X}_{i\ell n} / X_{i\ell n} - 1 \right| \times 100\%$.[25]

Discretization error is large even with many grid points in both the complements and substitutes case, regardless of the number of countries. Figure 4 graphs the discretization error against computational time for varying grid densities in both the complements and substitutes cases. Each line shows the precision-time frontier

[25] Since the support of productivity is unbounded, the only flows $X_{i\ell n}$ with zero value are those where $\gamma_{i\ell} = \infty$. In this case, the approximation $\tilde{X}_{i\ell n}$ is also zero, so we set the error to 0%.

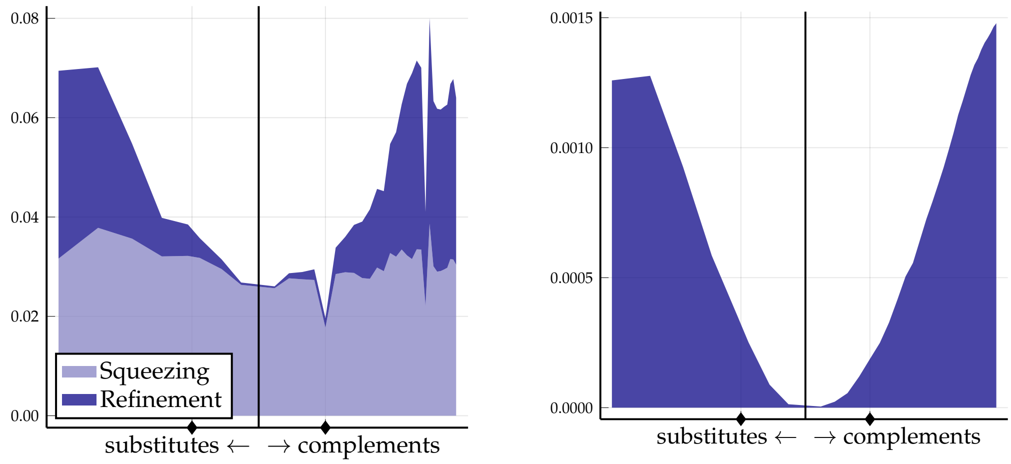
for a different number of countries. With 512 grid points and 8 countries, computation is fast, but the average discretization error exceeds 40% regardless of whether locations are complements or substitutes. Even with over 16000 grid points, the error remains near or above 20% in both cases. Doubling the number of grid points reduces the error by 5 to 10 percentage points but doubles the computation time.

Discretization error decreases with the number of countries in the complements case but shows no systematic pattern in the substitutes case, a difference that may partly reflect the different structure of the policy function in each. In the complements case, the policy function has a nesting structure: more productive firms produce in all locations chosen by less productive firms and possibly more.[26] Grid-based methods thus induce less error since they only miss marginal locations added between grid points, and this error tends to shrink as the number of countries grows. In the substitutes case, the policy function has no nesting structure: more productive firms may operate entirely different location sets, so grid-based methods may miss policy function switches between grid points entirely. Unlike the complements case, adding more countries does not systematically reduce this error.

Wide Applicability In a final exercise, we examine how computational time varies across the full range of our calibrations indexed by the degree of complementarity or substitutability among production locations. For each calibration, the left panel of Figure 5 shows the average time required to solve for the policy function across origin countries, using our policy function method. The figure also separates the computational time spent on squeezing versus the iterative cutoff search refinement. We mark the two benchmark calibrations with diamond symbols on the x-axis.

Across all degrees of complementarity and substitutability, our policy function method takes under 0.1 seconds on average to solve for the policy function. Computation is fastest near the vertical line where production locations are independent. As locations become more complementary or substitutable, computation time increases.

[26] Milgrom and Shannon [1994] show that the policy function exhibits a nesting structure when the objective function satisfies their single crossing and quasi-supermodularity conditions (the Online Appendix specifies the relationship between their conditions and SCD-C and SCD-T). Both conditions hold in our application, so that higher-productivity firms choose all locations selected by lower-productivity firms and possibly more: $z < z'$ implies $\mathcal{L}^\star(z) \subseteq \mathcal{L}^\star(z')$. The nesting result has been used in applied theory and empirical work, including by Antràs et al. [2017] in their analysis of firm-level sourcing decisions, and is reminiscent of the positive assortative patterns studied in Costinot [2009].



(A) Compute Time (s)

(B) Average Value of $\left|\overline{\mathcal{L}}^{(K)}(\cdot) \setminus \underline{\mathcal{L}}^{(K)}(\cdot)\right|$

FIGURE 5: DEGREE OF COMPLEMENTARITY OR SUBSTITUTABILITY AND COMPUTATIONAL PERFORMANCE

Note: This figure shows how computation time and the effectiveness of the policy function method vary with the degree of complementarity or substitutability across locations as measured by $\frac{\sigma-1}{\varepsilon-1} / \left(1 + \frac{\sigma-1}{\varepsilon-1}\right)$. The x-axis varies ε while holding $\sigma = 4$ fixed. The vertical line marks the case where locations are neither complements nor substitutes, with $\frac{\sigma-1}{\varepsilon-1} = 1$. We indicate the two benchmark calibrations with diamond symbols on the x-axis. The left panel decomposes the total computation time (in seconds) spent on generalized squeezing relative to the iterative cutoff search refinement. The right panel shows the average number of locations separating the upper and lower bounding set across all firms and countries after convergence of the generalized squeezing procedure. All computations were performed on an Apple M1 (2020) CPU.

Theorem 2 helps explain the patterns in the left panel. It establishes that the generalized squeezing method never takes more than $|L|$ applications, in line with the relatively flat squeezing time across all parameter combinations.

The right panel shows a proxy measure for the effectiveness of the squeezing procedure: the average number of locations separating the upper and lower bounding sets across all firms and countries after convergence of the generalized squeezing procedure. Few locations separate the bounding sets after convergence, though this number increases with the strength of complementarity or substitutability, consistent with the widening gap between squeezing and total time in the left panel.

4.2. Welfare Gains from Multinational Production

We examine how the welfare gains from MP vary with the degree of complementarity or substitutability among locations by comparing each calibrated equilibrium to its corresponding MP-autarky counterfactual in which firms are restricted to produce only in their headquarter country. Formally, the MP-autarky counterfactuals set bilateral MP costs to infinity, so that $\gamma_{i\ell} = \infty$ for all $i \neq \ell$.

Figure 6 reports the resulting welfare gains from MP across the full range of complementarity and substitutability. The light gray band shows the full cross-country range of effects and the dark gray band the interquartile range; the figure also highlights the welfare effects in selected countries. Two patterns stand out. First, at a given level of complementarity or substitutability, welfare gains from MP are uneven across countries. Second, this cross-country dispersion widens as locations become more complementary.

For a given degree of complementarity or substitutability, two main factors shape the differences in welfare gains across countries. The first is the cost of conducting MP, which differs across countries in the calibration: moving from MP autarky to the calibrated equilibrium thus represents a larger reduction in MP costs for some countries than for others. The Netherlands, for example, has its costs of doing MP lowered by more than most countries and therefore experiences larger welfare gains. The second is productivity. Countries can be productive both as headquarter locations—because of high headquarter productivity—and as production locations—because of low productivity-adjusted wages, low fixed costs, and good market access. Countries with these advantages benefit more from MP through the foreign expansion of their firms or by attracting foreign affiliates.

Moving from substitutability to complementarity in Figure 6, the distribution of welfare gains across countries fans out. As cross-location complementarity rises, welfare gains increase most in countries that serve as headquarter locations for highly productive firms, particularly those with low costs of conducting MP, while gains in economies without productive headquarters are dampened and can even turn negative. The welfare gains in the United States, for example, triple moving from strong substitutability to strong complementarity, whereas Lithuania and Romania experience losses with strong complementarity. With substitutability, by contrast, gains are more evenly distributed across countries.

This pattern arises because complementarity strengthens multinational scale economies, amplifying the marginal cost advantages of the most productive firms. As scale economies intensify, the most productive firms can leverage their affiliates to achieve even lower marginal costs and expand sales relative to less productive firms, so welfare gains concentrate in the countries where those firms are headquartered.

Countries that are less productive as headquarters and more productive as production locations see welfare losses, as foreign multinational set up local affiliates that crowd out domestic headquarters. When locations are substitutes, by contrast, production in one location cannibalizes production in others, limiting these scale effects and distributing gains more evenly across countries.

This concentration effect is reflected in the difference between the GDP-weighted average welfare gain across countries and the median welfare gain. The GDP-weighted average welfare gain is largest when complementarity is strongest, because scale economies at the most productive firms raise global output the most. The median welfare gain, however, is largest when substitutability is strongest, because gains are distributed more evenly and a broader set of countries benefits from MP.[27]

To further understand the sources of variation in welfare gains across countries, we decompose welfare gains following Arkolakis et al. [2012], adapted to our setting.[28] The welfare change in country i can be written as the sum of an openness term, a variety term, and an average productivity term:

$$\ln \frac{\hat{w}_i}{\hat{P}_i} = \underbrace{\ln \hat{\phi}_{iii}^{-\frac{1}{\sigma-1}}}_{\text{openness}} + \underbrace{\ln \hat{\tilde{z}}_i^{-\frac{\zeta}{\sigma-1}} + \ln \hat{M}_i^{\frac{1}{\sigma-1}}}_{\text{variety}} + \underbrace{\ln \hat{\tilde{z}}_i + \ln \left[\sum_{\mathcal{Z}_{i,t} \in \mathcal{T}_i} \lambda_{iii,t} s_{iii,t}^{1-\frac{\sigma-1}{\varepsilon-1}} \right]^{-\frac{1}{\sigma-1}}}_{\text{average productivity}} . \quad (9)$$

In this expression, $\hat{x} \equiv x/x'$ denotes the ratio of a variable in the baseline equilibrium to its value under MP autarky. The term ϕ_{iii} is the share of expenditure in country i on goods produced domestically by firms headquartered in i; it serves as an inverse measure of openness to trade and MP. The mass of entrants M_i and the productivity cutoff $\tilde{z}_i$ capture entry and selection effects in country i.

The policy function $\mathcal{L}_i^\star(\cdot)$ partitions firms into productivity intervals $\{\mathcal{Z}_{i,t}\}_t$, within which firms choose the same set of production locations. For firms in interval t, the term $s_{i\ell i,t}$ denotes the share of their domestic sales produced in country ℓ, so $s_{i\ell i,t}$ sums to 1 across production locations ℓ. The term $\lambda_{iii,t}$ denotes the share of total domestic sales, produced domestically, that are accounted for by firms in that

[27] The largest GDP-weighted average effect is 7.64%, occurring under the strongest complementarity, while the largest median welfare effect is 6.70%, occurring under the strongest substitutability.

[28] We derive this welfare formula in Appendix B under the assumption that, in the initial equilibrium, all active firms include the headquarter location in their optimal set of production locations. This assumption holds in all numerical exercises.

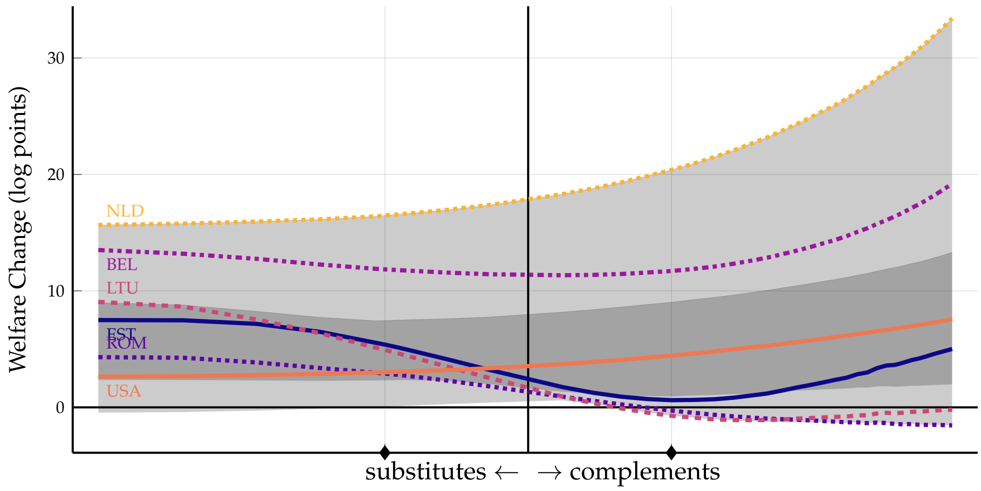


FIGURE 6: WELFARE GAINS FROM MULTINATIONAL PRODUCTION

Note: This figure shows the log point welfare change ($100 \times \ln\left(\hat{w}_i/\hat{P}_i\right)$) from moving from MP autarky to the calibrated economy, across model calibrations with different degrees of complementarity or substitutability among locations as measured by the ratio $\frac{\sigma-1}{\varepsilon-1}/\left(1+\frac{\sigma-1}{\varepsilon-1}\right)$. In this expression, $\hat{x} \equiv x/x'$ denotes the ratio of a variable in the baseline equilibrium to its value in MP autarky. The x-axis varies ε while holding $\sigma = 4$ fixed (its value in the baseline calibration). The vertical line indicates the case where locations are neither complements nor substitutes, with $\frac{\sigma-1}{\varepsilon-1} = 1$. The light gray band depicts the full range of welfare changes across all 32 countries in our calibration, while the dark gray band shows the interquartile range. The lines trace the welfare impact for select countries across the different degrees of complementarity or substitutability.

interval. The shares $\lambda_{iii,t}$ sum to 1 across intervals t. Together, these terms capture how MP reshapes the distribution of sales across firms and locations.

Figure 7 shows the welfare gains from MP for every country in our sample in pink, and their decomposition into the openness, variety, and average productivity channels in shades of blue. For readability, we show the welfare effects only for our benchmark degrees of complementarity and substitutability.

The openness channel is positive for all countries and reflects access to cheaper foreign production. Countries spend less on locally produced goods by domestic firms as they open to MP, effectively shifting their production possibility frontier outward. This effect is largest for two types of countries: economies that import large volumes and hence benefit from a global decrease in marginal costs through MP, and low-productivity economies that benefit from new, productive foreign affiliates producing locally, giving them access to foreign varieties at lower prices.

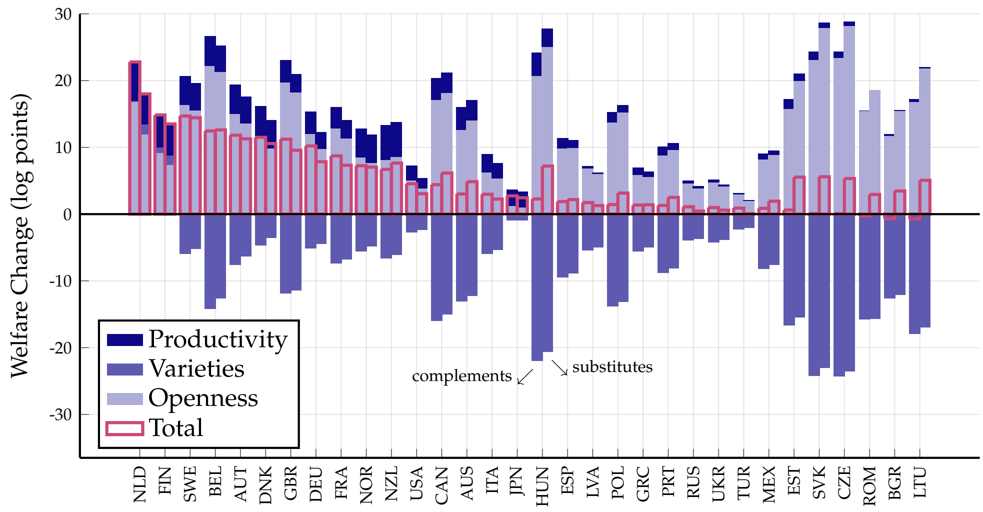


FIGURE 7: DECOMPOSING THE WELFARE GAINS FROM MULTINATIONAL PRODUCTION

Note: This figure shows the log point welfare change, $100 \times \ln\left(\hat{w}_i/\hat{P}_i\right)$, from moving from MP autarky to the calibrated economy as pink outlines. In addition, the figure decomposes the welfare changes into the contributions from changes in openness, the number of available varieties, and average productivity from equation (9). The left bars are for the complements calibration, the right bars for the substitutes calibration. The countries are ordered by the size of the total welfare effect in the complements calibration.

The variety channel reflects changes in the mass of domestic varieties. With complementarity, productive firms expand MP across locations, raising wages and survival cutoffs in host countries. Countries that mostly serve as hosts to the affiliates of foreign firms see their own domestic entrants crowded out, generating a negative variety effect. By contrast, in the Netherlands and Finland, countries where many multinationals are headquartered, global expansion raises profits, entry, and the mass of available local varieties. With substitutability, both effects are muted: multinationals operate fewer locations, earn lower profits, and crowd out fewer domestic firms in host countries.

The productivity channel is positive and captures changes in sales-weighted average productivity. In countries where multinationals are headquartered, the domestic market share of this class of productive firms expands relative to firms only producing domestically; in countries with few multinationals, this reallocation effect is small.[29] With complementarity, scale economies amplify this expansion and thus

[29]At the extreme, if no domestically-headquartered firm engages in MP, the distribution of domestic

the variation across countries of the productivity effect. Substitutability dampens the effect and its variation, since the competition among locations prevents multinational firms from expanding domestic production as much.

Together, these channels explain the variation in welfare gains from MP across countries and calibrations. Complementarity magnifies the advantages of highly productive and open economies and intensifies reallocation toward globally dominant firms. Substitutability limits this concentration, so gains are more even across countries. Thus, the strength and direction of interdependencies among production locations shape not only the aggregate gains from MP, but also their distribution across firms and countries.[30]

5. Conclusion

This paper studies combinatorial discrete choice problems, in which agents select items whose values depend on which others are chosen. Without structure on the objective function, finding the optimal decision set requires evaluating the payoffs of all possible combinations, a number that grows exponentially in the number of items, limiting the scope of quantitative models with such decision problems. We develop an exact solution method for such problems when the objective function satisfies a single crossing condition, which imparts structure on how each item's value changes as the chosen set expands. The method recovers the global optimum without exhaustive evaluation and extends to heterogeneous agents.

We apply the method to a general equilibrium model of multinational production in which heterogeneous firms choose international production locations subject to fixed costs. When locations are complements, scale economies concentrate production in the most productive firms, amplifying the welfare gains from multinational production in their headquarter countries while leading to small gains or even losses for countries that primarily host foreign affiliates. On the other hand, substitutability among locations limits this concentrating effect, so that gains are positive and more even across countries.

firms can be described by a representative firm as in the closed economy of Melitz [2003]. In this case, as a consequence of CES demand together with Pareto productivity, changes in MP costs cause no reallocations among firms, so that the productivity effect is zero. Competition for local labor instead lowers entry.

[30]In the Online Appendix, we examine how the welfare losses from MP autarky depend on the presence of fixed costs. Our benchmark calibration features fixed costs; without them, all firms produce in all countries and the CDCP disappears. Removing fixed costs lowers the concentration effect of MP. When we recalibrate the model without fixed costs to match the same data, the average welfare gains from MP are larger in both the complements and substitutes cases, in line with Tintelnot [2017].

Beyond multinational production, combinatorial discrete choice problems arise in a range of economic settings, such as sourcing decisions, network formation, technology adoption, and store location choices, yet are often simplified to preserve tractability at the cost of abstracting from economically relevant interdependencies. Our approach allows researchers to model such interdependencies and rich heterogeneity without sacrificing exactness or computational feasibility. Future research drawing on our tools could estimate the sign and strength of cross-choice interactions from micro data, analyze policies that alter complementarities or fixed costs, and quantify models with variable markups that respond to complementarity-driven scale economies.

References

Alonso Alfaro-Ureña, Juanma Castro-Vincenzi, Sebastián Fanelli, and Eduardo Morales. Firm export dynamics in interdependent markets. Technical report, National Bureau of Economic Research, 2024.

Vanessa Alviarez. Multinational production and comparative advantage. *Journal of International Economics*, 119:1–54, 2019.

Pol Antràs, Teresa C. Fort, and Felix Tintelnot. The margins of global sourcing: Theory and evidence from us firms. *American Economic Review*, 107(9):2514–64, 2017.

Pol Antràs, Evgenii Fadeev, Teresa C. Fort, and Felix Tintelnot. Exporting, global sourcing, and multinational activity: Theory and evidence from the united states. *The Review of Economics and Statistics*, pages 1–48, 04 2024a. ISSN 0034-6535. doi: 10.1162/rest_a_01450.

Pol Antràs, Evgenii Fadeev, Teresa C. Fort, and Felix Tintelnot. Export-platform fdi: Cannibalization or complementarity? *AEA Papers and Proceedings*, 114:130–35, May 2024b. doi: 10.1257/pandp.20241021.

Costas Arkolakis. Market penetration costs and the new consumers margin in international trade. *Journal of Political Economy*, 118(6):1151–1199, 2010.

Costas Arkolakis, Arnaud Costinot, and Andrés Rodríguez-Clare. New trade models, same old gains? *American Economic Review*, 102(1):94–130, 2012.

Costas Arkolakis, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple. Innovation and production in the global economy. *American Economic Review*, 108 (8):2128–73, 2018.

Costas Arkolakis, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare. The elusive pro-competitive effects of trade. *The Review of Economic Studies*, 86(1):46–80, 01 2019. ISSN 0034-6527. doi: 10.1093/restud/rdx075. URL https://doi.org/10.1093/restud/rdx075.

Costas Arkolakis, Fabian Eckert, and Rowan Shi. Combinatorial discrete choice: A quantitative model of multinational location decisions. Technical report, National Bureau of Economic Research, 2023.

Kristian Behrens, Giordano Mion, Yasusada Murata, and Jens Suedekum. Quantifying the Gap Between Equilibrium and Optimum under Monopolistic Competition. *The Quarterly Journal of Economics*, 135(4):2299–2360, 2020.

Steven T. Berry, James A. Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.

Endre Boros and Peter L. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1):155–225, 2002. ISSN 0166-218X. doi: https://doi.org/10.1016/S0166-218X(01)00341-9.

Christian Broda and David E. Weinstein. Globalization and the gains from variety. *The Quarterly Journal of Economics*, 121(2):541–585, 2006.

Maddalena Conte, Pierre Cotterlaz, Thierry Mayer, et al. *The CEPII Gravity Database*. 2023.

Arnaud Costinot. An elementary theory of comparative advantage. *Econometrica*, 77(4):1165–1192, 2009. doi: https://doi.org/10.3982/ECTA7636.

Sebastian Dyrda, Guangbin Hong, and Joseph B. Steinberg. A macroeconomic perspective on taxing multinational enterprises. *Journal of International Economics*, 152:104022, 2024. ISSN 0022-1996. doi: https://doi.org/10.1016/j.jinteco.2024.104022.

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

Pablo D Fajgelbaum, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar. State taxes and spatial misallocation. *The Review of Economic Studies*, 86(1):333–376, 2019.

Ying Fan and Chenyu Yang. Competition, product proliferation, and welfare: A study of the us smartphone market. *American Economic Journal: Microeconomics*, 12(2):99–134, 2020.

Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer. The next generation of the penn world table. *American Economic Review*, 105(10):3150–3182, 2015.

Matthew Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, 97(3):713–744, June 2007. doi: 10.1257/aer.97.3.713.

Pinelopi Koujianou Goldberg. Product differentiation and oligopoly in international markets: The case of the u.s. automobile industry. *Econometrica*, 63(4):891–951, 1995.

Keith Head and Thierry Mayer. Brands in motion: How frictions shape multinational production. *American Economic Review*, 109(9):3073–3124, 2019.

Keith Head, Thierry Mayer, Marc Melitz, and Chenying Yang. Industrial policies for multi-stage production: The battle for battery-powered vehicles. Working Paper 34884, National Bureau of Economic Research, February 2026. URL http://www.nber.org/papers/w34884.

Igal Hendel. Estimating multiple-discrete choice models: An application to computerization returns. *The Review of Economic Studies*, 66(2):423–446, 1999.

Jonas Hjort, Hannes Malmberg, and Todd Schoellman. The missing middle managers: Labor costs, firm structure, and development. Technical report, National Bureau of Economic Research, 2022.

Panle Jia. What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, 76(6):1263–1316, 2008.

Stephen C. Kleene. General recursive functions of natural numbers. *Mathematische Annalen*, 112:727–742, 12 1936.

Stephen C. Kleene. On notation for ordinal numbers. *Journal of Symbolic Logic*, 3(4):150–155, 1938.

Peter J Klenow and Jonathan L Willis. Real rigidities and nominal price changes. *Economica*, 83(331):443–472, 2016.

Jiří Klimeš. Fixed edge theorems for complete lattices. *Archivum Mathematicum*, 17(4):227–234, 1981.

Adrian Kulesza. Combinatorial discrete choice with deep reinforcement learning, 2024.

Nelson Lind and Natalia Ramondo. Trade with correlation. *American Economic Review*, 113(2):317–53, February 2023.

Daniel McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 1974.

Marc J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, 2003.

Paul Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, 2004.

Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62 (1):157–180, 1994.

Marc-Andreas Muendler and Sascha O. Becker. Margins of multinational labor substitution. *American Economic Review*, 100(5):1999—2030, 12 2010.

Robert A Pollak. Additive utility functions and linear engel curves. *The Review of Economic Studies*, 38(4):401–414, 1971.

Natalia Ramondo. A quantitative approach to multinational production. *Journal of International Economics*, 93(1):108–122, 2014.

Natalia Ramondo and Andrés Rodríguez-Clare. Trade, multinational production, and the gains from openness. *Journal of Political Economy*, 121(2):273–322, 2013.

Natalia Ramondo, Andrés Rodríguez-Clare, and Felix Tintelnot. Multinational production: Data and stylized facts. *American Economic Review*, 105(5):530–536, 2015.

JMC Santos Silva and Silvana Tenreyro. The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658, 2006.

Ina Simonovska. Income differences and prices of tradables: Insights from an online retailer. *The Review of Economic Studies*, 82(4):1612–1656, 2015.

Sebastian Sotelo. Practical aspects of implementing the multinomial pml estimator. *Ann Arbor: University of Michigan, mimeo*, 2019.

V. Stoltenberg-Hansen, I. Lindström, and E. R. Griffor. *Mathematical Theory of Domains*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1994.

Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309, 1955.

Feodora A. Teti. Missing tariffs. Technical Report 11590, CESifo Working Papers, 2024. Feodora Teti's Global Tariff Database (v_beta1-2024-12).

Felix Tintelnot. Global production with export platforms. *The Quarterly Journal of Economics*, 132(1):157–209, 2017.

Kenneth E. Train, Daniel McFadden, and Moshe Ben-Akiva. The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The RAND Journal of Economics*, 18(1), 1987.

A. Appendix A: Methods for Policy Function Refinement

In this section, we introduce two methods of refining the policy function after the generalized squeezing procedure has converged. Generalized squeezing produces two reduced bounding functions that partition the type space into intervals, on each of which the lower and upper bounding functions take constant values $[\underline{\mathcal{L}}_t, \overline{\mathcal{L}}_t]$. Where the bounding functions coincide, the optimal decision set equals their common value. Where they do not, further refinement is required. We focus on solving for the policy function over one such interval $\mathcal{Z}_t$ with $\underline{\mathcal{L}}_t \subset \overline{\mathcal{L}}_t$; the full policy function is recovered by applying a refinement method to each such interval in turn.

The first refinement method, generalized branching, requires SCD-C and SCD-T. The second, iterative cutoff search, requires the stronger condition of strong SCD-T and is the method we use in our quantitative application. The Online Appendix contains all proofs.

A.1. The Generalized Branching Procedure

We extend the branching logic from Section 1 to heterogeneous agent settings. On an interval $\mathcal{Z}_t$ where $\underline{\mathcal{L}}_t \subset \overline{\mathcal{L}}_t$, the generalized branching step selects one location $\ell \in \overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t$ and creates two *branches*: one that imposes ℓ is included in the decision sets and the other that imposes ℓ is excluded.

Each branch corresponds to the problem of solving for the policy function that maps each type $z \in \mathcal{Z}_t$ to the conditionally optimal decision set $\mathcal{L}$ maximizing the modified objective function $\tilde{f} : \mathcal{P}\left(\left(\overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t\right) \setminus \{\ell\}\right) \times \mathcal{Z}_t \to \mathbb{R}$ where:

$$\begin{aligned} \tilde{f}(\mathcal{L}, z) &\equiv f(\mathcal{L} \cup \underline{\mathcal{L}}_t \cup \{\ell\}, z) && \text{on the branch that includes } \ell, \\ \tilde{f}(\mathcal{L}, z) &\equiv f(\mathcal{L} \cup \underline{\mathcal{L}}_t, z) && \text{on the branch that excludes } \ell. \end{aligned}$$

The generalized branching step applies the generalized squeezing procedure to each restricted CDCP, yielding a pair of reduced bounding functions for each branch. It then translates these bounding functions for the restricted CDCPs into bounding functions for the original CDCP by adding all items in $\underline{\mathcal{L}}_t$ to both bounding functions on each branch. On the branch that includes ℓ, the item ℓ is also added to both bounding functions. The output of the generalized branching step is thus a pair of conditional bounding functions for the original CDCP on each branch which each induce a new partition on $\mathcal{Z}_t$.

On type intervals in this new partition for which the bounding functions on a branch coincide, the branch terminates, having identified a conditionally optimal

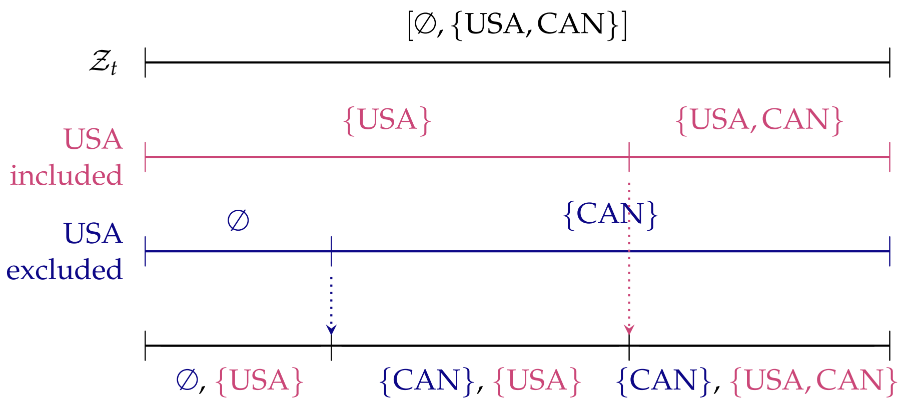


FIGURE 8: THE GENERALIZED BRANCHING PROCEDURE

Note: The figure illustrates an example outcome of the generalized branching procedure, in an example where $\underline{\mathcal{L}}_t = \emptyset$ and $\overline{\mathcal{L}}_t = \{\text{USA}, \text{CAN}\}$. The branching step selects $\ell = \text{USA}$ and creates two branches: one which presumes USA is in the optimal set and the other which presumes the opposite. Convergence on a single branch occurs when the generalized squeezing procedure returns a conditionally optimal policy function. The final output of the full recursive generalized branching procedure is the collection of all conditionally optimal policy functions.

policy function; on other intervals, the generalized branching step is applied recursively to each sub-interval where the bounding functions differ. We refer to the recursive application of the generalized branching step until all branches terminate on all intervals as the "generalized branching procedure." Because each step strictly reduces the number of locations separating the upper and lower bounding functions, the procedure branches a finite number of times.

To obtain the global policy function on $\mathcal{Z}_t$, we combine the results across all terminal branches. Each terminal branch yields a conditionally optimal policy function, which is piecewise constant but may change value at different productivity types across branches. We intersect these policy functions to partition $\mathcal{Z}_t$ into subintervals of types on which none of the conditionally optimal policy functions change value. On each such subinterval, we collect the candidate decision sets from the conditionally optimal policy functions. For each type $z \in \mathcal{Z}_t$, the global policy function chooses the decision set that yields the highest value among these candidate decision sets.

Figure 8 illustrates this refinement on an interval $\mathcal{Z}_t$ where the converged bounding functions are $\underline{\mathcal{L}}_t = \emptyset$ and $\overline{\mathcal{L}}_t = \{\text{USA}, \text{CAN}\}$. Branching on the USA location creates two branches: one that includes the USA in the optimal set, in pink, and one that excludes it, in blue. In the example, applying generalized squeezing on each branch produces a conditionally optimal policy function that maps each type to an optimal decision set, shown in the two middle lines. The bottom line collects the candidate optimal sets for each element of the type space partition. The optimal policy function is found by choosing, for each type $z \in \mathcal{Z}_t$, the profit-maximizing decision set among the conditionally optimal decision sets.

A.2. Iterative Cutoff Search

Iterative cutoff search is an alternative refinement method that relies on the additional structure that a stronger version of SCD-T induces on the policy function.

Definition (Strong SCD-T)**.** For any pair of decision sets $(\mathcal{L}, \mathcal{L}')$ and types $z < z'$,

$$f(\mathcal{L}, z) - f(\mathcal{L}', z) \geq 0 \qquad \Rightarrow \qquad f(\mathcal{L}, z') - f(\mathcal{L}', z') > 0$$

Strong SCD-T requires that, for any pair of decision sets, there is at most one type that receives the same payoff from the two sets. When the objective function satisfies strong SCD-T, the policy function does not change value and then revert to a previous value as the type increases: if two types share the same optimal decision set, all intermediate types share it as well. The policy function is thus a sequence of sub-intervals, each with a distinct optimal decision set.

Consider an interval $\mathcal{Z}_t$ from generalized squeezing with end points $z < z'$. If the optimal decision sets on the two end points coincide, so that $\mathcal{L}^\star(z) = \mathcal{L}^\star(z')$, then this decision set must also be optimal for all types within the interval.

If the optimal decision sets on the two end points do not coincide, we apply iterative cutoff search. It pins down the policy function two adjacent sub-intervals at a time by narrowing the interval until there are only two unique values of the policy function. It then finds the cutoff type where the optimal decision set switches, using the fact that both decision sets must be optimal and deliver the same value at the cutoff.

Given an interval where $\mathcal{L}^\star(z) \neq \mathcal{L}^\star(z')$, strong SCD-T guarantees a unique type $\bar{z} \in [z, z']$ where the two optimal decision sets of the end points have the same value. If either of the sets is also optimal at $\bar{z}$, then the policy function must switch from $\mathcal{L}^\star(z)$ to $\mathcal{L}^\star(z')$ at $\bar{z}$. The interval is therefore resolved into two sub-intervals, each with constant optimal decision sets.[31]

If instead $\mathcal{L}^\star(\bar{z})$ differs from both end point sets, then there are at least three sub-intervals within $[z, z']$, since three distinct values of the policy function have been

[31] As we discuss in Appendix B, the model of our quantitative application admits a closed-form expression for the indifferent type.

identified. Iterative cutoff search narrows its search to $[z, \overline{z}]$ and iterates again. Once it succeeds in finding two adjacent sub-intervals of the policy function, as described, it returns to iterate on the remainder of the interval.

Definition (Iterative cutoff search)**.** Consider the interval $\left[z^{(0)}, z'^{(0)}\right)$. Iterate as follows.

1. Identify the type $\overline{z}^{(k)}$ that is indifferent between $\mathcal{L}^\star\left(z^{(k)}\right)$ and $\mathcal{L}^\star\left(z'^{(k)}\right)$. Solve for the optimal decision set at this type, $\mathcal{L}^\star\left(\overline{z}^{(k)}\right)$, over the reduced domain implied by $\left[\underline{\mathcal{L}}_t, \overline{\mathcal{L}}_t\right]$.
2. Update as follows.
   a) If $\mathcal{L}^\star\left(\overline{z}^{(k)}\right)$ coincides with either $\mathcal{L}^\star\left(z^{(k)}\right)$ or $\mathcal{L}^\star\left(z'^{(k)}\right)$, then update the policy function

   $$\mathcal{L}^\star(z) = \begin{cases} \mathcal{L}^\star\left(z^{(k)}\right) & \text{for } z \in \left[z^{(k)}, \overline{z}^{(k)}\right) \\ \mathcal{L}^\star\left(z'^{(k)}\right) & \text{for } z \in \left[\overline{z}^{(k)}, z'^{(k)}\right) \end{cases}.$$

   If $z'^{(k)} = z'^{(0)}$, the policy function has been solved for the full interval. Terminate iteration. Otherwise, set $z^{(k+1)} = z'^{(k)}$ and $z'^{(k+1)} = z'^{(0)}$.

   b) Otherwise, set $z^{(k+1)} = z^{(k)}$ and $z'^{(k+1)} = \overline{z}^{(k)}$.

Computing $\mathcal{L}^\star(\overline{z})$ at the candidate cutoff types requires solving a CDCP. We implement this step using the squeezing and branching methods from the main paper. Thus, while SCD-C is not required for iterative cutoff search, we use it to compute the optimal decision sets in practice.

B. Appendix B: Model Derivations

In this section, we establish the SCD-C and SCD-T conditions in the firm problem of Section 2 and derive the welfare expression in equation (9). In the Online Appendix, we establish the single crossing differences conditions in a broader class of frameworks.

B.1. Establishing SCD-C and SCD-T

Let $\mathcal{L} \subset \mathcal{L}'$. The marginal value of location ℓ to the set $\mathcal{L}'$ in terms of its marginal value to $\mathcal{L}$ is:

$$\begin{aligned} D_\ell \pi_i \left(\mathcal{L}', z\right) = D_\ell \pi_i \left(\mathcal{L}, z\right) & \\ + \left(\frac{\sigma-1}{\varepsilon-1} - 1\right) \frac{\sigma-1}{\varepsilon-1} \sum_{n \in L} \frac{1}{\sigma} X_n \left(\frac{\sigma-1}{\sigma} P_n z\right)^{\sigma-1} & \\ \times \int_0^{\zeta_{i\ell n}^{1-\varepsilon}} \int_{\Theta_{in}(\mathcal{L}\setminus\{\ell\})}^{\Theta_{in}(\mathcal{L}'\setminus\{\ell\})} \left[t+u\right]^{\frac{\sigma-1}{\varepsilon-1}-2} \mathrm{d}u \mathrm{d}t & \end{aligned}$$

where we denote $\Theta_{in}\left(\mathcal{L}\right) \equiv \sum_{\ell \in \mathcal{L}} \zeta_{i\ell n}^{1-\varepsilon}$. Since $\mathcal{L}'$ nests $\mathcal{L}$, $\Theta_{in}\left(\mathcal{L}'\right) \geq \Theta_{in}\left(\mathcal{L}\right)$ for all markets n and the integrals are positive.

As the marginal value shows, the profit function satisfies the SCD-C condition. When the ratio $(\sigma-1)/(\varepsilon-1)$ is above 1, the marginal value of location ℓ in $\mathcal{L}'$ is positive whenever its marginal value in $\mathcal{L}$ is positive. In this case, the profit function satisfies SCD-C from below. On the other hand, when the ratio is below 1, the marginal value of ℓ is negative in $\mathcal{L}'$ whenever it is negative in $\mathcal{L}$, so that the profit function satisfies SCD-C from above.

Now, let $\{\mathcal{L}, \mathcal{L}'\}$ be any pair of production location sets. The difference in profits associated with each of these decision sets can be written:

$$\begin{aligned} \pi_i\left(\mathcal{L}', z\right) - \pi_i\left(\mathcal{L}, z\right) = z^{\sigma-1} \sum_{n \in L} \frac{1}{\sigma} X_n \left(\frac{\sigma-1}{\sigma} P_n\right)^{\sigma-1} \left[\Theta_{in}\left(\mathcal{L}'\right)^{\frac{\sigma-1}{\varepsilon-1}} - \Theta_{in}\left(\mathcal{L}\right)^{\frac{\sigma-1}{\varepsilon-1}}\right] & \\ - \sum_{\ell \in L} w_\ell f_{i\ell} \left(\mathbb{1}\left[\ell \in \mathcal{L}'\right] - \mathbb{1}\left[\ell \in \mathcal{L}\right]\right) & \end{aligned}$$

Since $\sigma > 1$, this difference is strictly monotonic in the productivity z, so there is at most one productivity at which the firm is indifferent between the two production location sets.[32] Then, the profit function satisfies strong SCD-T, and, in turn, SCD-T. The type indifferent between the two production sets can be expressed in closed form.

B.2. Counterfactual Welfare

The share of expenditure in n that is captured by goods produced in ℓ by firms originating from i, denoted $\phi_{i\ell n}$, integrates over the behavior of all firms. It simplifies

[32] Strong SCD-T is violated only in the knife-edge case where two decision sets exist such that the total variable profit improvement (summed across markets n) is zero *and* the difference in total fixed cost payments (summed across locations ℓ) is also zero.

to:

$$\phi_{i\ell n} = M_i \left(\frac{\sigma-1}{\sigma} P_n\right)^{\sigma-1} \times \sum_{t=1}^{T} \mathbb{1}_{i\ell,t} \xi_{i\ell n}^{1-\varepsilon} \Theta_{in} \left(\mathcal{L}_{i,t}\right)^{\frac{\sigma-1}{\varepsilon-1}-1} \left[\left(z_{i,t+1}\right)^{\sigma-1-\zeta} - \left(z_{i,t}\right)^{\sigma-1-\zeta}\right] \tilde{\zeta}$$

where $\tilde{\zeta}$ is a constant of integration. We split the productivity range of active firms into the intervals along which the policy function is constant, indexed by $t \in \{1, \ldots, T\}$, denoting with $z_{i,t}$ the left end point of each interval and setting $z_{i,T+1} = \infty$. Within each interval, let $\mathcal{L}_{i,t}$ be the optimal set and $\mathbb{1}_{i\ell,t} \equiv \mathbb{1}\left[\ell \in \mathcal{L}_{i,t}\right]$ indicate whether location ℓ is in the optimal set.

In what follows, we assume that fundamentals are such that all active firms establish production in their country of origin i, so that $\mathbb{1}_{ii,t} = 1$ for all t. This assumption holds in all our calibrations.

For MP autarky, we set $\gamma'_{i\ell} = \infty$ for all $i \neq \ell$, so that active firms only produce domestically. Then, there is only one interval, $[\tilde{z}'_i, \infty)$, and the optimal location set along this interval is $\{i\}$. The counterfactual change in ϕ_{iii} simplifies to:

$$\frac{\phi'_{iii}}{\phi_{iii}} = \left(\frac{M'_i}{M_i}\right)\left(\frac{P'_i}{P_i}\right)^{\sigma-1}\left(\frac{\xi'_{iii}}{\xi_{iii}}\right)^{1-\sigma}\left(\frac{\tilde{z}'_i}{\tilde{z}_i}\right)^{\sigma-1-\zeta} \sum_t \lambda_{iii,t} s_{iii,t}^{1-\frac{\sigma-1}{\varepsilon-1}},$$

where

$$s_{i\ell i,t} \equiv \frac{\xi_{i\ell i}^{1-\varepsilon}}{\sum_{\ell' \in \mathcal{L}_{i,t}} \xi_{i\ell' i}^{1-\varepsilon}},$$

$$\lambda_{iii,t} \equiv \frac{s_{iii,t}\Theta_{ii}\left(\mathcal{L}_{i,t}\right)^{\frac{\sigma-1}{\varepsilon-1}}\left[\left(z_{i,t}\right)^{\sigma-1-\zeta} - \left(z_{i,t+1}\right)^{\sigma-1-\zeta}\right]}{\sum_{t'=1}^{T} s_{iii,t'}\Theta_{ii}\left(\mathcal{L}_{i,t'}\right)^{\frac{\sigma-1}{\varepsilon-1}}\left[\left(z_{i,t'}\right)^{\sigma-1-\zeta} - \left(z_{i,t'+1}\right)^{\sigma-1-\zeta}\right]}.$$

Using $\xi'_{iii}/\xi_{iii} = w'_i/w_i$, we arrive at equation (9).

C. Appendix C: Computational Implementation of Our Solution Methods

We implement all solution methods necessary for the quantitative exercises in the Julia package accompanying this paper, available at `https://github.com/rowanxshi/CDCP.jl`. The package contains the squeezing and branching procedures to solve the CDCP at a specific firm productivity. It also contains the generalized squeezing procedure and iterative cutoff search to solve for the policy function across a range of firm productivity values. Our replication package, which calls the Julia solution package, implements all quantitative exercises with the multinational model. This section discusses how our methods enter the quantitative exercises of Sections 3 and 4.

C.1. Calibration of GE Model

The first application of our solution methods is to calibrate the model in general equilibrium. Calibration requires solving for the policy functions for each origin country, $\{\mathcal{L}_i^\star(\cdot)\}_i$, at every new parameter guess. We thus embed the policy function method within the calibration routine.

As Table 1 lays out, we first set the parameters $\{\sigma, \varepsilon, \zeta\}$. Our calibration finds model parameters such that the data values of $\{w_\ell, M_\ell, H_\ell\}_\ell$ are an equilibrium.

We start with an initial guess for the market size aggregate $\left\{P_n^{(0)}\right\}_n$, the fundamentals $\left\{A_\ell^{(0)}, \underline{z}_\ell^{(0)}, f_\ell^{(0)}\right\}_\ell$, and the bilateral cost parameterization $\left\{\overline{\tau}_\ell^{(0)}, \overline{\gamma}_\ell^{(0)}, \overline{\nu}_\ell^{(0)}\right\}_\ell$ with $\left\{\kappa_\tau^{v(0)}, \kappa_\gamma^{v(0)}, \kappa_\nu^{v(0)}\right\}_v$. We then iterate as follows.

1. Solve the policy functions $\{\mathcal{L}_i^\star(\cdot)\}_i$ of the firm's CDCP from equation (2) using the policy function method. The policy function also determines $\tilde{z}_i$ as the lowest type to operate at least one location, satisfying equation (3).
2. Given $\{\mathcal{L}_i^\star(\cdot)\}_i$, compute trade, MP, and affiliate flows in the model; estimate the PPML gravity equation for each (see the Online Appendix for details).
3. Use deviations from aggregate conditions and moments to update the guess:
   - update $\left\{P_n^{(k)}\right\}_n$ using equation (5);
   - update $\left\{A_\ell^{(k)}, \underline{z}_\ell^{(k)}, f_\ell^{(k)}\right\}_\ell$ using respectively deviations from equation (6), the gap between model and data share of total foreign MP that is attributable to firms headquartered in each country, and the gap between model and data survival rate in each country;
   - update $\left\{\overline{\tau}_\ell^{(k)}, \overline{\gamma}_\ell^{(k)}, \overline{\nu}_\ell^{(k)}\right\}_\ell$ using the gap between model and data own shares of trade (domestic sales), MP (domestic production), and affiliates (domestic production locations); and

- update $\left\{\kappa_{\tau}^{v(k)}, \kappa_{\gamma}^{v(k)}, \kappa_{\nu}^{v(k)}\right\}_{v}$ using the gap between the model and data coefficients of the PPML regression.

The iteration converges to (near) zero on all deviations. The routine thus calibrates the model while enforcing GE conditions at the same time. After convergence, we invert equation (4) to recover f_i^e.

C.2. Time Trials and GE Counterfactuals

Section 4 begins with a subsection benchmarking the computational performance of our methods using the calibrated models. In Table 2 and Figure 4, we compare the squeezing method, discretized on a grid over the productivity range, against the policy function method. To do so, we use each approach to solve the CDCP across the full range of firm productivity values, holding fixed all model parameters, fundamentals, and aggregates. In Figure 5, for each value of the ratio $(\sigma - 1)/(\varepsilon - 1)$, we calibrate the model as described above and apply the policy function method. None of these figures requires computing the general equilibrium of the model.

To compute the welfare effects from multinational production requires solving the model in general equilibrium. To do so, given a set of parameters and fundamentals, we use an iterative routine that solves for the policy functions, $\{\mathcal{L}_i^{\star}(\cdot)\}_i$, at each new guess for aggregates.

We start with an initial guess for the aggregates $\left\{P_\ell^{(0)}, w_\ell^{(0)}, M_\ell^{(0)}\right\}_\ell$, then iterate as follows.

1. Given aggregates $\left\{P_\ell^{(k)}, w_\ell^{(k)}, M_\ell^{(k)}\right\}_\ell$, set $X_\ell = w_\ell H_\ell$ for all ℓ to satisfy equation (7).
2. Solve the policy functions $\{\mathcal{L}_i^{\star}(\cdot)\}_i$ of the firm's CDCP from equation (2) using the policy function method. The policy function also determines $\tilde{z}_i$ as the lowest type to operate at least one location, satisfying equation (3).
3. Given the policy functions $\{\mathcal{L}_i^{\star}(\cdot)\}_i$, use deviations from aggregate conditions to update the guess:
   - update $\left\{P_\ell^{(k)}\right\}_\ell$ using equation (5);
   - update $\left\{w_\ell^{(k)}\right\}_\ell$ using equation (6); and
   - update $\left\{M_\ell^{(k)}\right\}_\ell$ using equation (4).

In the counterfactual of Section 4, solving for the policy function becomes trivial. Firms open no foreign locations, so they compare $\mathcal{L} = \emptyset$, which represents being inactive, against $\mathcal{L} = \{i\}$, which represents producing domestically. Solving for the policy function then simplifies to identifying the cutoffs $\{\tilde{z}_i\}_i$, where the fixed cost of setting up the domestic location perfectly offsets the variable profits that the firm would earn.

SUPPLEMENTAL ONLINE APPENDIX

Combinatorial Discrete Choice: Theory and Application to Multinational Production

Costas Arkolakis
Yale University
NBER
costas.arkolakis@yale.edu

Fabian Eckert
UCSD
fpe@ucsd.edu

Rowan Shi
TMU
rowanxshi@torontomu.ca

FOR ONLINE PUBLICATION ONLY

D. Data and Calibration

This section discusses our data construction and calibration in more detail.

D.1. Data

We describe the data inputs in our calibration and their sources. The replication package documents the raw data, processing and imputation procedures, and detailed download instructions.

Trade, Foreign Affiliate Sales, and Foreign Affiliate Counts We obtain data on bilateral trade flows, foreign affiliate sales, and foreign affiliate counts from the dataset compiled by Alviarez [2019]. This dataset combines four major sources: OECD International Direct Investment Statistics and the Statistics on Measuring Globalization; Eurostat Foreign Affiliate Statistics; Bureau of Economic Analysis (BEA) public data; and Bureau van Dijk's Orbis. All values are averages over 2003–2012; we use the same period for all other data sources in our calibration.

The data cover 32 countries and nine sectors. We retain all countries but aggregate across manufacturing industries. For each country pair, we observe bilateral trade flows. We construct domestic sales by summing a country's total export sales and subtracting them from the total sales of the sector in the country to obtain sales to the domestic market. Using these home sales, we can then construct home trade shares.

For each origin-destination pair, we observe total sales of foreign affiliates—for example, total sales of Canadian-owned firms located in Germany. The data do not report the final destination of these sales: we do not observe how much Canadian affiliates in Germany sell to Greece. We compute sales by domestic firms as total sales in a country minus sales by foreign affiliates, from which we construct the complete matrix of multinational sales volume for each origin and destination country.

The data contain counts of foreign affiliates by country and sector but not counts of domestic firms. We supplement with UNIDO data on total firm counts by country, averaged over 2003–2012 (see below). We compute the number of domestic firms as total firms minus foreign affiliates, which we interpret as headquarters in the model. From these counts, we construct the full matrix of multinational production affiliate counts for each origin and destination country.

Gravity Variables We obtain standard gravity variables from the CEPII database [see Conte et al., 2023] with minor modifications. For distance, we use the simple distance between countries' most populated cities, measured in kilometers. For colonial ties, we combine CEPII's "colonial sibling" dummy (indicating a common past colonizer) and "colonial dependence" dummy (indicating a direct colonial relationship) into a single "colonial relationship" indicator. We also include dummies for shared borders and common official language. The bilateral trade agreement dummy is the only gravity variable that varies over our sample period; we set it to its median value over 2003–2012.

We also use GDP per capita, total population, and total employment from the Penn World Table in the data package by Alviarez [2019].

Bilateral Tariffs We obtain bilateral tariff data from Teti [2024], which constructs a comprehensive panel of ad valorem tariffs from the UNCTAD TRAINS database. The underlying data contain MFN (most favored nation) rates and preferential rates for over 5000 HS6 product categories. For each product-country pair, Teti [2024] takes the minimum of MFN and preferential rates, drops non-ad valorem tariffs, and aggregates to broader sector groupings. We use the agriculture/non-agriculture file and retain only non-agricultural tariffs, consistent with our focus on manufacturing. We average tariff rates over 2003–2012 to match our other data sources.

Firm Counts Data on the number of manufacturing establishments come from the UNIDO Industrial Statistics Database (INDSTAT). We use total manufacturing establishments for the 32 countries in our sample, averaged over 2003–2012. For country-year observations with missing data, we impute values using a regression of log establishments on log GDP, log population, and country and year fixed effects.

Firm Survival Rates Data on firm survival rates come from the OECD Structural and Demographic Business Statistics database. The survival rate measures the percentage of firms born in year t that remain active in year $t + n$, for horizons of one to five years. We use data on firms (those with employees, excluding self-employed) in the manufacturing sector, aggregated across all firm sizes. For country-year observations with missing data, we impute values using a regression of log survival rates on log GDP interacted with horizon, log population interacted with horizon, and year and employer-type fixed effects. We average across 2003–2012 to obtain a single survival rate per country. In our baseline calibration, we use the one-year survival rate, as it corresponds most closely to the concept of firms that pay an entry cost to learn their productivity but exit before producing positive quantities.

D.2. Calibration

In this section, we provide more detail on the calibration of our model.

Bilateral Costs of Trade, MP, and Affiliates We separately estimate the following empirical gravity equation for trade flows, inward MP sales, and inward affiliate stocks, indexed by x.

$$y_{ij}^{x} = \exp\left(\alpha^{x} + \sum_{v\in\{d,\text{COL},\text{COM},\text{BOR}\}} \beta_{v}^{x} v_{ij} + \delta_{x}' X_{ij} + \varkappa_{i} + \zeta_{j}\right) + \epsilon_{ij}^{x} \tag{10}$$

The gravity variables, indexed by v, are the log distance d_{ij} between countries i and j, and dummies for colonial relations (COL), common language (COM) and common borders (BOR). The vector X_{ij} contains additional gravity controls, in particular bilateral tariffs and a free trade agreement dummy. The terms $\varkappa_i$ and ζ_j are origin and destination specific fixed effects.

Table 3 presents results for the gravity variables coefficients that we target in calibration, omitting untargeted coefficients for conciseness. We estimate equations via Poisson Pseudo Maximum Likelihood [see Silva and Tenreyro, 2006] as well as OLS, where the estimated elasticities are in line with those of similar regressions in Ramondo et al. [2015]. In computing manufacturing tariffs, it is necessary to decide whether to use raw averages of the tariffs of all goods in manufacturing or weight in some way. For robustness, we report results using both unweighted and weighted tariffs.

We specify the following functional forms of the bilateral trade, MP, and fixed costs:

$$\begin{aligned}
\log \tau_{\ell n} &= \overline{\tau}_{n} \times \mathbb{1}\left[\ell \neq n\right] + \sum_{v\in\{d,\text{COL},\text{COM},\text{BOR}\}} \kappa_{v}^{\tau} v_{\ell n} + \log\left(1 + t_{\ell n}\right) \\
\log \gamma_{i\ell} &= \overline{\gamma}_{\ell} \times \mathbb{1}[i \neq \ell] + \sum_{v\in\{d,\text{COL},\text{COM},\text{BOR}\}} \kappa_{v}^{\gamma} v_{i\ell} \\
\log \nu_{i\ell} &= \overline{\nu}_{\ell} \times \mathbb{1}[i \neq \ell] + \sum_{v\in\{d,\text{COL},\text{COM},\text{BOR}\}} \kappa_{v}^{\nu} v_{i\ell}
\end{aligned} \tag{11}$$

where $t_{\ell n}$ are tariffs and v indexes the same set of gravity variables as in the regression (10). The $\{\overline{\tau}_n, \overline{\gamma}_\ell, \overline{\nu}_\ell\}$ components represent the costs of doing an activity across borders versus within borders. In our calibration procedure, we estimate the destination-specific components $\{\overline{\tau}_n, \overline{\gamma}_\ell, \overline{\nu}_\ell\}$ by targeting the own-shares

TABLE 3: TRADE, MP, AND FOREIGN AFFILIATE GRAVITY IN THE DATA

	Unweighted Tariffs			Weighted Tariffs		
	(1) Trade	(2) MP	(3) Affiliates	(4) Trade	(5) MP	(6) Affiliates
Log Distance	-.7201***	-.3275***	-.6923***	-.7532***	-.3822***	-.6974***
	(.05532)	(.09722)	(.08641)	(.05143)	(.105)	(.08654)
Colony	.04607	-.02342	.3119**	.05978	-.07851	.2874*
	(.1273)	(.1327)	(.1483)	(.1138)	(.1278)	(.1474)
Contiguity	.4558***	.3279*	.3943***	.4634***	.3303**	.3906***
	(.07158)	(.1721)	(.1001)	(.06841)	(.1657)	(.1005)
Language	.144	.5467***	.6298***	.1342	.4939***	.6017***
	(.1032)	(.1563)	(.1774)	(.09662)	(.1554)	(.1789)
Observations	992	992	992	992	992	992

(A) Estimation with PPML

	Unweighted Tariffs			Weighted Tariffs		
	(1) Trade	(2) MP	(3) Affiliates	(4) Trade	(5) MP	(6) Affiliates
Log Distance	-1.131***	-.8242***	-.798***	-1.144***	-.821***	-.8015***
	(.05193)	(.1214)	(.07376)	(.05168)	(.1218)	(.07407)
Colony	.7346***	.9477***	.6667***	.7029***	.9546***	.6577***
	(.1087)	(.2431)	(.1481)	(.1082)	(.2442)	(.1489)
Contiguity	.3209***	.7195***	.3895***	.3218***	.721***	.3965***
	(.09246)	(.1957)	(.1185)	(.09175)	(.1957)	(.1186)
Language	-.01175	.597**	.2354	-.001395	.5941**	.234
	(.1356)	(.2846)	(.1733)	(.1347)	(.2847)	(.1735)
Observations	992	707	710	992	707	710

(B) Estimation with OLS

Note: The table presents the coefficients from estimating gravity equations. The outcome variable differs across the columns: bilateral manufacturing trade flows, bilateral multinational production sales, and bilateral foreign affiliate stocks. The standard gravity controls serve as explanatory variables. All estimating equations also include origin and destination fixed effects and additional controls for bilateral tariffs and a regional trade agreement dummy. Tariffs are averages across all manufacturing goods, either unweighted or weighted by the global trade shares of each good. The specifications exclude the diagonal entries of the respective flow matrix. Robust standard errors are in parentheses. We denote different levels of significance as follows: *** Significant at 1 percent level, ** Significant at 5 percent level, and * Significant at 10 percent level.

of each activity, and the gravity-variable-specific elasticities by targeting the estimated coefficients on the corresponding gravity variables in Table 3 using PPML and unweighted tariffs. Table 9a reports the estimated elasticities.

Figure 9b shows histograms of our calibrated trade costs, MP costs, and the bilateral component of fixed costs; we exclude the diagonal entries of all cost matrices from the histogram since they are normalized to 1. Our estimated trade costs are substantial, similar to prior estimates from studies featuring trade and multinational production [e.g. Ramondo and Rodríguez-Clare, 2013].

In contrast, our estimated MP costs are small compared to previous studies such as Ramondo and Rodríguez-Clare [2013] or Arkolakis et al. [2018]. These differences arise from the fact that we allow for fixed costs in addition to MP cost. We use affiliate count data to separate fixed costs from MP cost, while previous studies that only use MP sales and trade data cannot separate these two costs. In the data, there are few MP affiliates, but they account for a large sales volume in their host countries; to match these patterns, we estimate large fixed costs and therefore smaller MP costs. The presence of economically significant fixed costs is in line with Hjort et al. [2022], which finds that, in real terms, labor compensation of middle management is an important component of the cost of doing multinational business abroad and does not vary much across MNE locations.

Estimated Fundamentals In the left panels of Figure 10, we plot our estimates of the headquarter productivity shifter $\underline{z}_i$ of the firm Pareto distribution against the location productivity shifter A_ℓ. Recall that $\underline{z}_i$ is identified by the share of foreign affiliate sales attributable to firms headquartered in country i, while A_ℓ is chosen to match countries' level of GDP per capita. The United States has the highest headquarter productivity in our dataset. Relatively developed economies with little MNE activity, such as Greece and Spain, lie below the US and to the right of the 45 degree line that defines the US comparative advantage. On the opposite side of the 45 degree line and close to the US lie developed countries with relatively high MNE activity such as Netherlands, Germany, and Finland.

In the right panel of Figure 10, we plot the entry cost and the base component of the fixed cost for each country. In our data, one-year survival rates range from 77% to 92%, so that the ratio between the base component of the fixed cost and the entry cost varies little across countries.

Figure 11 graphs the trade shares, inward MP sales shares, and inward foreign affiliate shares generated by the calibrated models against the same objects in the data. The model provides a good fit, especially for larger shares. The fit for larger shares is better since the targeted PPML specification in Table 3 is in levels, thereby putting relatively more weight on larger countries [see Sotelo, 2019, for a discussion]. The model-generated data produces exactly the same coefficient estimates as in Table 3.

Model Fit Figure 12 presents our calibrated model's performance on an untargeted moment: the US sales premium of multinational firms based in the US. We calculate in the model the average sales among groups of US-based firms with presence in different minimum numbers of foreign locations, normalized by the average sales of non-MNE firms based in the US. Figure 12 shows these MNE sales premia computed in our benchmark calibrated models and the empirical premia documented by Antràs et al. [2024a]. The model premia closely mirror the empirical premia, reflecting that more productive firms broadly have both more foreign affiliate locations and higher sales. For example, MNEs (any firm conducting production in at least two countries) are 30 times larger than non-MNEs in the US in the model and 40 times in the data.

Alternative Calibration without Fixed Costs To understand the importance of fixed costs in shaping counterfactual responses of the economy to economic shocks, we compare the welfare gains from multinational production in the model calibrated with and without fixed costs.

Eliminating fixed costs simplifies computation by making the CDCP trivial [see, e.g., Ramondo and Rodríguez-Clare, 2013, Ramondo, 2014, Arkolakis et al., 2018, Fajgelbaum et al., 2019]. All firms survive and set up production locations in every country $\ell \in L$, only facing the intensive margin problem of choosing how much to produce in each. The profit function collapses to

$$\pi_i(z) = \sum_n \frac{1}{\sigma} q_n\left(p_{in}(L,z)\right) p_{in}(L,z)\,.$$

We recalibrate the model, setting the fixed costs to zero, so that $f_i = 0$ for all origins i. We follow the same procedure as with the full model, but drop as calibration targets the survival rate of firms and the coefficients in the third column of Table 3.

Figure 13 reports the welfare impact of MP autarky, comparing the effects in the benchmark calibrations with fixed costs against the alternative calibrations without fixed costs. For most countries, the gains from MP are larger without fixed costs because all firms produce in all countries, so MP lowers every firm's marginal cost symmetrically. There is no concentration effect: the relative sizes of firms remain the same. By contrast, with fixed costs, relatively unproductive firms are unable to survive, and the variety loss dampens the welfare gains.

E. The Mathematics of CDCPs

In this section, we provide formal proofs for statements in the paper.

E.1. Definitions and Existing Results

Definition (Poset)**.** A poset (partially-ordered set) $(P, \leq)$ is a set together with a partial ordering that is:

1. reflexive: for all $x \in P$, $x \leq x$;
2. antisymmetric: for any pair with $x \leq y$ and $y \leq x$, it must be that $x = y$; and
3. transitive: for any elements with $x \leq y$ and $y \leq z$, it must be that $x \leq z$.

The dual poset $(P, \leq_D)$ is the set P together with the ordering that is defined $x \leq_D y$ iff $y \leq x$.

Definition (Order-preserving and order-reversing)**.** An endomap on the poset $(P, \leq)$, $\Phi : P \to P$, is order-preserving if, given $x < y$, $\Phi(x) \leq \Phi(y)$. It is order-reversing if, given $x < y$, $\Phi(y) \leq \Phi(x)$. If the mapping is either order-preserving or order-reversing, it is monotonic.

Definition (Directed-complete)**.** A poset $(P, \leq)$ is directed-complete if, for all directed subsets $D \subseteq P$, $\sup D$ exists. The subset D is a *directed subset* if it is non-empty and, for every pair $x, y \in D$, there is a $z \in D$ with $x \leq z$ and $y \leq z$.

Definition (Scott continuity)**.** A function between two posets $f : (P, \leq_P) \to (Q, \leq_Q)$ is Scott-continuous if, for every directed subset $D \subseteq P$ whose supremum exists, the image of the supremum is the supremum of the image: $f(\sup\{D\}) = \sup\{f(x) \mid x \in D\}$.

Theorem (Kleene)**.** *Given a directed-complete partial order $(D, \leq)$ with a least element x and Scott-continuous endomap $f : D \to D$, f has a least fixed point, which is* $\sup\{f^n(x) \mid n \in \mathbb{N}\}$.

Definition (Lattice)**.** A poset $(L, \leq)$ is a lattice if, for any pair of elements $x, y \in L$, there is a unique supremum $\sup\{x, y\}$ and infimum $\inf\{x, y\}$ with respect to $\leq$. The lattice is *complete* if, for any subset $S \subseteq L$, there is a unique supremum $\sup S$ and infimum $\inf S$. A sublattice $(L', \leq)$ is a subset of points $L' \subseteq L$ that is itself a lattice.

Theorem (Tarski [1955])**.** *Given a complete lattice $(L, \leq)$ and an order-preserving endomap $f : L \to L$, the set of fixed points of f forms a complete lattice.*

Theorem (Klimeš [1981]). *Given a complete lattice $(L, \leq)$ and an order-reversing endomap $f : L \to L$, there is a least element u of L so that $(u, f(u))$ is a fixed edge of f. There is similarly a greatest element v with $(v, f(v))$ a fixed edge of f. Moreover, $v = f(u)$.*

E.2. Proofs: Solving CDCPs

The following results establish not only that the set of fixed points of Φ is nonempty and contains a global maximizer of f, but also a strategy to find *all* the maximizers of f, starting from the set of fixed points of Φ. They require no structure on f.

Theorem. *Let $\Xi \equiv \{\mathcal{L} \mid \mathcal{L} = \Phi(\mathcal{L})\}$ be the set of fixed points of Φ. Every decision set $\mathcal{L}^\star \in \arg\max_{\mathcal{L} \in \mathcal{P}(L)} f(\mathcal{L})$ that maximizes the objective function can be recovered from a decision set $\mathcal{L}' \in \arg\max_{\mathcal{L} \in \Xi} f(\mathcal{L})$ by removing a sequence of zero marginal value elements.*

Proof. Select an arbitrary maximizer $\mathcal{L}^\star \in \arg\max_{\mathcal{L} \in \mathcal{P}(L)} f(\mathcal{L})$. Then, no single deviation can strictly improve the value, so:

$$\begin{cases} D_\ell f(\mathcal{L}^\star) > 0 & \Rightarrow \ell \in \mathcal{L}^\star \\ D_\ell f(\mathcal{L}^\star) < 0 & \Rightarrow \ell \notin \mathcal{L}^\star \end{cases}.$$

If there is no element $\ell \notin \mathcal{L}^\star$ where $D_\ell f(\mathcal{L}^\star) = 0$, then $\mathcal{L}^\star$ is itself a fixed point of Φ so that $\mathcal{L}^\star \in \arg\max_{\mathcal{L} \in \Xi} f(\mathcal{L})$.

Suppose there is at least one element $\ell \notin \mathcal{L}^\star$ where $D_\ell f(\mathcal{L}^\star) = 0$, so that $\mathcal{L}^\star \notin \Xi$. We construct $\mathcal{L}'$ as follows. Let $\mathcal{L}_0 = \mathcal{L}^\star$ and iterate

$$\mathcal{L}_k = \mathcal{L}_{k-1} \cup \{\ell_k\} \qquad , \qquad 0 = D_{\ell_k} f(\mathcal{L}_{k-1}), \ \ell_k \notin \mathcal{L}_{k-1}.$$

Terminate once there is no element $\ell_k \notin \mathcal{L}_{k-1}$ with $D_{\ell_k} f(\mathcal{L}_{k-1}) = 0$. Iteration terminates, since each iteration either adds an element to $\mathcal{L}_{k-1}$ or terminates, and there are $|L|$ total elements. Let it terminate at $\mathcal{L}_K$.[33]

Since each decision set $\mathcal{L}_k$ is a series of zero-marginal-value additions to $\mathcal{L}^\star$, it attains the same value:

$$f(\mathcal{L}_k) = f(\mathcal{L}^\star) + \sum_{j=1}^{k} D_{\ell_j} f\left(\mathcal{L}_{j-1}\right) = f(\mathcal{L}^\star).$$

[33] This definition is ambiguous because it does not specify which element ℓ_k to choose at each iteration k if there are multiple candidates. However, we show that an arbitrary outcome of this sequence is in $\arg\max_{\mathcal{L} \in \Xi} f(\mathcal{L})$, which is sufficient for the proof.

so that $\mathcal{L}_k \in \arg\max_{\mathcal{L}\in\mathcal{P}(L)} f(\mathcal{L})$ for each k. The terminal decision set $\mathcal{L}_K$ is also a fixed point of Φ by construction. Thus, $\mathcal{L}_K \in \arg\max_{\mathcal{L}\in\Xi} f(\mathcal{L})$.

Then, $\mathcal{L}^\star$ can be recovered from $\mathcal{L}_K \in \arg\max_{\mathcal{L}\in\Xi} f(\mathcal{L})$ by iteratively removing the zero-marginal-value elements $\{\ell_k, \ldots, \ell_1\}$. □

Corollary. *Let $\Xi \equiv \{\mathcal{L} \mid \mathcal{L} = \Phi(\mathcal{L})\}$ be the set of fixed points of Φ. Then,*

1. *There is a global maximizer of f within Ξ, so*
$$\arg\max_{\mathcal{L}\in\Xi} f(\mathcal{L}) \subseteq \arg\max_{\mathcal{L}\in\mathcal{P}(L)} f(\mathcal{L}).$$
2. *The global maximizer of f is unique iff both there is a unique $\mathcal{L}^\star \in \arg\max_{\mathcal{L}\in\Xi} f(\mathcal{L})$, and $D_\ell f(\mathcal{L}^\star) \neq 0$ for all $\ell \in L$.*

Proof. Follows from the previous theorem. □

Single Crossing Differences in Choices We first show that SCD-C from above and below are necessary and sufficient for Φ to be order-reversing and order-preserving, respectively.

Proposition. *Consider the objective function $f : \mathcal{P}(L) \to \mathbb{R}$.*

1. *Quasi-supermodularity of f is sufficient for SCD-C from below; quasi-submodularity is sufficient for SCD-C from above.*
2. *If L is finite, then f is supermodular iff it has weakly increasing marginal values; submodular iff it has weakly decreasing marginal values.*

Proof. We show the statements for quasi-submodularity, submodularity, and SCD-C from above. Similar arguments follow for quasi-supermodularity, supermodularity, and SCD-C from below.

1. Suppose f satisfies quasi-submodularity: i.e. for any pair of elements $x, y \in \mathcal{P}(L)$,
$$f(x \cup y) - f(y) \geq 0 \quad \Rightarrow \quad f(x) - f(x \cap y) \geq 0$$
and let $\mathcal{L} \subset \mathcal{L}' \in \mathcal{P}(L)$, $\ell \in L$ with $D_\ell f(\mathcal{L}') \geq 0$. We show that $D_\ell(\mathcal{L}) \geq 0$. Let $\mathcal{J} \equiv \mathcal{L} \cup \{\ell\}$ and $\mathcal{K} \equiv \mathcal{L}' \setminus \{\ell\}$. Then,
$$D_\ell f(\mathcal{L}') = f(\mathcal{L}' \cup \{\ell\}) - f(\mathcal{L}' \setminus \{\ell\}) = f(\mathcal{J} \cup \mathcal{K}) - f(\mathcal{K}) \geq 0$$
$$\Rightarrow f(\mathcal{J}) - f(\mathcal{J} \cap \mathcal{K}) \geq 0$$

where the last line follows from quasi-submodularity. Then, it immediately follows that $D_\ell(\mathcal{L}) \geq 0$.

2. Suppose f is submodular: i.e. for any pair of elements $x, y \in \mathcal{P}(L)$,

$$f(x \cup y) + f(x \cap y) \leq f(x) + f(y).$$

We show it has weakly decreasing marginal values. Let $\mathcal{L} \subset \mathcal{L}'$ and consider the marginal value of the element ℓ. Let $x = \mathcal{L}' \setminus \{\ell\}$ and $y = \mathcal{L} \cup \{\ell\}$. Then,

$$\begin{aligned} D_\ell f(\mathcal{L}') - D_\ell f(\mathcal{L}) &= \left[f\left(\mathcal{L}' \cup \{\ell\}\right) - f\left(\mathcal{L}' \setminus \{\ell\}\right)\right] - \left[f(\mathcal{L} \cup \{\ell\}) - f(\mathcal{L} \setminus \{\ell\})\right] \\ &= f(x \cup y) - f(x) - f(y) + f(x \cap y) \leq 0 \end{aligned}$$

where the inequality follows from submodularity. Then, f has weakly decreasing marginal values.

Suppose f has weakly decreasing marginal values and let x and y be arbitrary elements of $\mathcal{P}(L)$. Let $\mathcal{L} \equiv x \cap y$, $\mathcal{J} \equiv y \setminus x$, and $\mathcal{K} \equiv x \setminus y$. The sets $\mathcal{J}$, $\mathcal{K}$, and $\mathcal{L}$ are disjoint and

$$\begin{aligned} &f(x \cup y) + f(x \cap y) - f(x) - f(y) \\ &= \left[f(\mathcal{K} \cup \mathcal{L} \cup \mathcal{J}) - f(\mathcal{K} \cup \mathcal{L})\right] - \left[f(\mathcal{L} \cup \mathcal{J}) - f(\mathcal{L})\right] \end{aligned}$$

so it suffices for submodularity to show this difference is non-positive. We use induction on $|\mathcal{J}|$.

Suppose $|\mathcal{J}| = 1$. WLOG, let $\mathcal{J} \equiv \{\ell\}$. Then,

$$\begin{aligned} &\left[f(\mathcal{K} \cup \mathcal{L} \cup \{\ell\}) - f(\mathcal{K} \cup \mathcal{L})\right] - \left[f(\mathcal{L} \cup \{\ell\}) - f(\mathcal{L})\right] \\ &= D_\ell(\mathcal{K} \cup \mathcal{L}) - D_\ell(\mathcal{L}) \end{aligned}$$

since $\ell \notin \{\mathcal{K} \cup \mathcal{L}\}$. By weakly decreasing marginal values, we establish the base case. Suppose weakly decreasing marginal values implies submodularity as long as $|\mathcal{J}| = k$ and consider the case of $|\mathcal{J}| = k + 1 > 1$. Select an element $\ell \in \mathcal{J}$ and define $\tilde{\mathcal{J}} \equiv \mathcal{J} \setminus \{\ell\}$ so that $\left|\tilde{\mathcal{J}}\right| = k > 0$. Then,

$$\begin{aligned} &\left[f(x \cup y) - f(x)\right] - \left[f(y) - f(x \cap y)\right] \\ &= \left[f\left(\mathcal{K} \cup \mathcal{L} \cup \tilde{\mathcal{J}} \cup \{\ell\}\right) - f(\mathcal{K} \cup \mathcal{L})\right] - \left[f\left(\mathcal{L} \cup \tilde{\mathcal{J}} \cup \{\ell\}\right) - f(\mathcal{L})\right] \\ &= \left[f\left(\mathcal{K} \cup \mathcal{L} \cup \tilde{\mathcal{J}} \cup \{\ell\}\right) - f\left(\mathcal{K} \cup \mathcal{L} \cup \tilde{\mathcal{J}}\right) + f\left(\mathcal{K} \cup \mathcal{L} \cup \tilde{\mathcal{J}}\right) - f(\mathcal{K} \cup \mathcal{L})\right] \\ &\quad - \left[f\left(\mathcal{L} \cup \tilde{\mathcal{J}} \cup \{\ell\}\right) - f\left(\mathcal{L} \cup \tilde{\mathcal{J}}\right) + f\left(\mathcal{L} \cup \tilde{\mathcal{J}}\right) - f(\mathcal{L})\right] \\ &= \left[f\left(\mathcal{K} \cup \mathcal{L} \cup \tilde{\mathcal{J}} \cup \{\ell\}\right) - f\left(\mathcal{K} \cup \mathcal{L} \cup \tilde{\mathcal{J}}\right)\right] - \left[f\left(\mathcal{L} \cup \tilde{\mathcal{J}} \cup \{\ell\}\right) - f\left(\mathcal{L} \cup \tilde{\mathcal{J}}\right)\right] \\ &\quad + \left[f\left(\mathcal{K} \cup \mathcal{L} \cup \tilde{\mathcal{J}}\right) - f(\mathcal{K} \cup \mathcal{L})\right] - \left[f\left(\mathcal{L} \cup \tilde{\mathcal{J}}\right) - f(\mathcal{L})\right] \end{aligned}$$

where difference is non-positive by weakly decreasing marginal values and the inductive assumption.

□

Proposition. *The mapping Φ defined in Definition 4 is*

1. *order-reversing iff the underlying objective function f obeys SCD-C from above; and*
2. *order-preserving iff the underlying objective function f obeys SCD-C from below.*

Proof. We show the first statement. Let $\mathcal{L} \subset \mathcal{L}'$ be two arbitrary nested decision sets.

Start with the converse. Suppose f obeys SCD-C from above. If $\Phi(\mathcal{L}')$ is empty, then it is contained in $\Phi(\mathcal{L})$ trivially; so let $\ell \in \Phi(\mathcal{L}')$ be an arbitrary element. Then, by definition of Φ,

$$D_\ell f\left(\mathcal{L}'\right) \geq 0 \qquad \overset{\text{SCD-C above}}{\Rightarrow} \qquad D_\ell f\left(\mathcal{L}\right) \geq 0$$

so $\ell \in \Phi(\mathcal{L})$. Then, $\Phi(\mathcal{L}') \subseteq \Phi(\mathcal{L})$ and Φ is order-reversing.

Now consider the forward direction. Let ℓ be an arbitrary element so that $D_\ell f\left(\mathcal{L}'\right) \geq 0$. If no such element exists, then SCD-C from above holds vacuously, so suppose at least one such ℓ exists. Then, by definition, $\ell \in \Phi(\mathcal{L}') \subseteq \Phi(\mathcal{L})$ since Φ is order-reversing. Then, by definition of Φ, it must be that $D_\ell f\left(\mathcal{L}\right) \geq 0$.

A similar argument holds for SCD-C from below. □

The Squeezing Procedure We now formally prove Theorem 1.

Proof. Let $\mathcal{L}^\star$ be an arbitrary maximizer in $\{\mathcal{L} \mid \mathcal{L} = \Phi(\mathcal{L})\}$. Since Φ is monotonic,

$$\left[\underline{\mathcal{L}}^{(k+1)}, \overline{\mathcal{L}}^{(k+1)}\right] = \begin{cases} \left[\Phi\left(\overline{\mathcal{L}}^{(k)}\right), \Phi\left(\underline{\mathcal{L}}^{(k)}\right)\right] & \text{with SCD-C from above} \\ \left[\Phi\left(\underline{\mathcal{L}}^{(k)}\right), \Phi\left(\overline{\mathcal{L}}^{(k)}\right)\right] & \text{with SCD-C from below} \end{cases}$$

and thus as long as $\underline{\mathcal{L}}^{(k)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k)}$, the monotonicity of Φ guarantees $\underline{\mathcal{L}}^{(k+1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k+1)}$ because $\Phi(\mathcal{L}^\star) = \mathcal{L}^\star$.

We now show that the bounding sets weakly tighten each iteration using induction on the iteration. Start with $k = 0$. Then, $\emptyset = \underline{\mathcal{L}}^{(0)} \subseteq \underline{\mathcal{L}}^{(1)} \subseteq \overline{\mathcal{L}}^{(1)} \subseteq \overline{\mathcal{L}}^{(0)} = L$

trivially. Assume $\underline{\mathcal{L}}^{(k-1)} \subseteq \underline{\mathcal{L}}^{(k)} \subseteq \overline{\mathcal{L}}^{(k)} \subseteq \overline{\mathcal{L}}^{(k-1)}$. Applying Φ to each decision set and using monotonicity,

$$\begin{cases} \Phi\left(\overline{\mathcal{L}}^{(k-1)}\right) \subseteq \Phi\left(\overline{\mathcal{L}}^{(k)}\right) \subseteq \Phi\left(\underline{\mathcal{L}}^{(k)}\right) \subseteq \Phi\left(\underline{\mathcal{L}}^{(k-1)}\right) & \text{with SCD-C from above} \\ \Phi\left(\underline{\mathcal{L}}^{(k-1)}\right) \subseteq \Phi\left(\underline{\mathcal{L}}^{(k)}\right) \subseteq \Phi\left(\overline{\mathcal{L}}^{(k)}\right) \subseteq \Phi\left(\overline{\mathcal{L}}^{(k-1)}\right) & \text{with SCD-C from below} \end{cases}$$

$$\Rightarrow \underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$$

from the definition of S.

Finally, each iteration of the squeezing step must add at least one additional item to the lower bounding set or exclude at least one additional item from the upper bounding set which can occur a maximum of $|L|$ times. □

Corollary. *If f satisfies SCD-C, the squeezing step S itself is an order-preserving mapping.*

Proof. Define the partial order $\leq_{[\,]}$ over the collection of bounding sets so that $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] \leq_{[\,]} \left[\underline{\mathcal{L}}', \overline{\mathcal{L}}'\right]$ iff both $\underline{\mathcal{L}} \subseteq \underline{\mathcal{L}}'$ and $\overline{\mathcal{L}}' \subseteq \overline{\mathcal{L}}$, i.e. order the bounding sets by tightness. We first show that $\leq_{[\,]}$ is a valid order relation:

1. it is reflexive: $\underline{\mathcal{L}} \subseteq \underline{\mathcal{L}}$ and $\overline{\mathcal{L}} \subseteq \overline{\mathcal{L}}$ so $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] \leq_{[\,]} \left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]$;
2. it is antisymmetric: if $\underline{\mathcal{L}} \subseteq \underline{\mathcal{L}}'$ and $\overline{\mathcal{L}} \subseteq \overline{\mathcal{L}}'$ but also $\underline{\mathcal{L}}' \subseteq \underline{\mathcal{L}}$ and $\overline{\mathcal{L}}' \subseteq \overline{\mathcal{L}}$, then $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] = \left[\underline{\mathcal{L}}', \overline{\mathcal{L}}'\right]$; and
3. it is transitive: if $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] \leq_{[\,]} \left[\underline{\mathcal{L}}', \overline{\mathcal{L}}'\right]$ and $\left[\underline{\mathcal{L}}', \overline{\mathcal{L}}'\right] \leq_{[\,]} \left[\underline{\mathcal{L}}'', \overline{\mathcal{L}}''\right]$ then $\underline{\mathcal{L}} \subseteq \underline{\mathcal{L}}' \subseteq \underline{\mathcal{L}}''$ and $\overline{\mathcal{L}}'' \subseteq \overline{\mathcal{L}}' \subseteq \overline{\mathcal{L}}$ so $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] \leq_{[\,]} \left[\underline{\mathcal{L}}'', \overline{\mathcal{L}}''\right]$.

We prove the corollary for SCD-C from above. The proof for SCD-C from below follows similarly.

Let $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] \leq_{[\,]} \left[\underline{\mathcal{L}}', \overline{\mathcal{L}}'\right]$ be two arbitrary ordered bounding sets. Since Φ is order-reversing, $S\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right) = \left[\Phi\left(\overline{\mathcal{L}}\right), \Phi\left(\underline{\mathcal{L}}\right)\right]$ is a pair of bounding sets and similarly for $\left[\underline{\mathcal{L}}', \overline{\mathcal{L}}'\right]$. Since $\overline{\mathcal{L}}' \subseteq \overline{\mathcal{L}}$ and Φ order-reversing, we have that $\Phi\left(\overline{\mathcal{L}}\right) \subseteq \Phi\left(\overline{\mathcal{L}}'\right)$. Similarly, $\underline{\mathcal{L}} \subseteq \underline{\mathcal{L}}'$ implies that $\Phi\left(\underline{\mathcal{L}}'\right) \subseteq \Phi\left(\underline{\mathcal{L}}\right)$. Thus,

$$\left[\Phi\left(\overline{\mathcal{L}}\right), \Phi\left(\underline{\mathcal{L}}\right)\right] \leq_{[\,]} \left[\Phi\left(\overline{\mathcal{L}}'\right), \Phi\left(\underline{\mathcal{L}}'\right)\right]$$

which completes the proof.

Note that the bounding sets $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]$ define an implicit sublattice on $(\mathcal{P}(L), \subseteq)$. Thus, letting $\mathcal{B}$ be the set of sublattices of $(\mathcal{P}(L), \subseteq)$, the mapping S is order-preserving on the lattice $(\mathcal{B}, \supseteq)$. $\square$

We now prove that the squeezing procedure identifies the smallest and largest fixed point of Φ when f satisfies SCD-C from below and the extreme fixed edge when f satisfies SCD-C from above.

Lemma. *In a finite poset, every directed subset contains its supremum and the poset is directed-complete.*

Proof. Let $(P, \leq)$ be a finite poset and $D \subseteq P$ a directed subset. The statement is clearly true if $|D| = 1$. We proceed using induction on $|D|$. Suppose $\sup D$ exists and is in D when $|D| = k$.

Suppose $|D| = k + 1$. Select two distinct elements $x, x' \in D$ and let $\overline{x}$ be an element with $x \leq \overline{x}$ and $x' \leq \overline{x}$. At least one of x or x' is distinct from $\overline{x}$. WLOG, let $x \neq \overline{x}$. Then, $D \setminus \{x\}$ is also directed, since $\overline{x} \in D \setminus \{x\}$.

Let $y = \sup(D \setminus \{x\})$ which exists and is contained in $D \setminus \{x\}$ by inductive assumption. Then, since D is directed, there must be an element $z \in D$ where $y \leq z$ and $x \leq z$. If $z \notin D \setminus \{x\}$, then $z = x$. In that case, $y \leq x$, so x is an upper bound of D. It is also the least upper bound; for any weakly smaller upper bound $u \leq x$, we have $x \leq u$ since u is an upper bound. Since $\leq$ is antisymmetric, $u = x$. Then, $x = \sup D \in D$. On the other hand, if $z \in D \setminus \{x\}$, it must be that $z = y$ since y is the supremum of $D \setminus \{x\}$ so $z \leq y$. Then, $x \leq y$ so $y = \sup D \in D$. $\square$

Lemma. *Let $(P, \leq)$ be a finite poset and $f : (P, \leq) \to (P, \leq)$ be an order-preserving endomap. Then, f is Scott-continuous.*

Proof. Let $D \subseteq P$ be a directed subset. By the previous lemma, its supremum exists and is in D. Since f is order-preserving, $f(x) \leq f(\sup D)$ for all $x \in D$, so $f(\sup D)$ is an upper bound on the image of D; but also $\sup D \in D$, so $f(\sup D) \in \{f(x) \mid x \in D\}$ implies that $f(\sup D)$ is the least upper bound. $\square$

Corollary. *When f satisfies SCD-C, so that $\left[\underline{\mathcal{L}}^{(K)}, \overline{\mathcal{L}}^{(K)}\right] = S^K([\emptyset, L])$ is a fixed point by Theorem 1, then*

1. if f satisfies SCD-C from below, $\underline{\mathcal{L}}^{(K)}$ and $\overline{\mathcal{L}}^{(K)}$ are the smallest and largest fixed points of Φ respectively; and

2. if f satisfies SCD-C from above, the pair $\underline{\mathcal{L}}^{(K)}$ and $\overline{\mathcal{L}}^{(K)}$ is the extreme fixed edge of Φ.

Proof. We show each statement in turn.

1. In this case, Φ is order-preserving. By the previous lemmas, $(\mathcal{P}(L), \subseteq)$ is directed-complete and Φ is Scott-continuous. Directly applying Kleene's fixed point theorem, $\underline{\mathcal{L}}^{(K)} = \Phi^K(\emptyset)$ is the smallest fixed point of Φ.

   Now consider the dual poset $(\mathcal{P}(L), \supseteq)$. The least element in the dual poset is L, since $L \supseteq \mathcal{L}$ for all $\mathcal{L} \in \mathcal{P}(L)$. Define Φ_D as an endomap on the dual poset with $\Phi_D(\mathcal{L}) \equiv \Phi(\mathcal{L})$. Then, Φ_D is order-preserving: for all $\mathcal{L} \supset \mathcal{L}'$, we have $\Phi_D(\mathcal{L}) \supseteq \Phi_D(\mathcal{L}')$. By an identical argument as above, it is Scott-continuous, so Kleene's fixed point theorem implies $\overline{\mathcal{L}}^{(K)} = \Phi_D^K(L)$ is the smallest fixed point on Φ_D. It is thus the largest fixed point of Φ.

2. Define

   $$E \equiv \left\{\mathcal{L} \mid \exists \mathcal{L}' \in \mathcal{P}(L) \text{ where } \mathcal{L} = \Phi\left(\mathcal{L}'\right) \text{ and } \mathcal{L}' = \Phi(\mathcal{L})\right\}$$

   as the collection of sets which are part of a fixed edge of Φ. Then, E is also the set of fixed points of Φ^2: for every $\mathcal{L} \in E$, $\Phi(\Phi(\mathcal{L})) = \Phi(\mathcal{L}') = \mathcal{L}$ so $\mathcal{L}$ is a fixed point of Φ^2; and for every fixed point of Φ^2, $\mathcal{L} = \Phi^2(\mathcal{L})$, $\Phi(\Phi(\mathcal{L})) = \mathcal{L}$ so $\mathcal{L} \in E$ with $\mathcal{L}' = \Phi(\mathcal{L})$.

   Because Φ^2 is order-preserving, by identical arguments we have that $\underline{\mathcal{L}}^{(K)} = \Phi^{2K}(\emptyset)$ is the smallest fixed point of Φ^2, i.e. the smallest element of E. Then, $\overline{\mathcal{L}}^{(K)} = \Phi\left(\underline{\mathcal{L}}^{(K)}\right)$ and $\underline{\mathcal{L}}^{(K)} = \Phi\left(\overline{\mathcal{L}}^{(K)}\right)$, so $\overline{\mathcal{L}}^{(K)} \in E$. We show it is the largest element of E to complete the proof. Let $\mathcal{L}^{\text{sup}} \in E$ be the greatest element. Then, $\overline{\mathcal{L}}^{(K)} \subseteq \mathcal{L}^{\text{sup}}$. Define $\mathcal{L}' = \Phi(\mathcal{L}^{\text{sup}})$ and note that it is also in E. Then, $\mathcal{L}' = \Phi(\mathcal{L}^{\text{sup}}) \subseteq \Phi\left(\overline{\mathcal{L}}^{(K)}\right) = \underline{\mathcal{L}}^{(K)}$ since Φ is order-reversing. Then, $\mathcal{L}' = \underline{\mathcal{L}}^{(K)}$ since $\underline{\mathcal{L}}^{(K)}$ is the smallest element of E, so $\Phi(\mathcal{L}') = \mathcal{L}^{\text{sup}} = \overline{\mathcal{L}}^{(K)}$ is the greatest element.

□

The Branching Procedure In what follows, we denote $\mathcal{L} \cup \left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] \equiv \left[\mathcal{L} \cup \underline{\mathcal{L}}, \mathcal{L} \cup \overline{\mathcal{L}}\right]$ and write $\Phi(\cdot; f)$, $S(\cdot; f)$, or $S^K(\cdot; f)$ to specify the objective function *and domain* used with each mapping.

Definition (Branching step)**.** Given bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ and an element $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$, define $\tilde{L} \equiv (\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}) \setminus \{\ell\}$ and the conditional objective function

$$\tilde{f}(\cdot; \underline{\mathcal{L}}) : \mathcal{P}(\tilde{L}) \to \mathbb{R} \qquad , \qquad \tilde{f}(\mathcal{L}; \underline{\mathcal{L}}) \equiv f(\mathcal{L} \cup \underline{\mathcal{L}}).$$

Then, the branching step is defined as

$$B_\ell([\underline{\mathcal{L}}, \overline{\mathcal{L}}]) \equiv \Big\{ (\underline{\mathcal{L}} \cup \{\ell\}) \cup S^K([\emptyset, \tilde{L}]; \tilde{f}(\cdot; \underline{\mathcal{L}} \cup \{\ell\})), \\ \underline{\mathcal{L}} \cup S^K([\emptyset, \tilde{L}]; \tilde{f}(\cdot; \underline{\mathcal{L}})) \Big\}$$

Note that the branching step is well-defined since, as in Theorem 1, the squeezing procedure converges when applied iteratively to the domain bounding sets.

Lemma. *Given a fixed point $\mathcal{L}'$ of Φ and bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ with $\underline{\mathcal{L}} \subseteq \mathcal{L} \subseteq \overline{\mathcal{L}}$, the set $\mathcal{L}' \setminus \underline{\mathcal{L}}$ is a fixed point of $\Phi(\cdot; \tilde{f}(\cdot; \underline{\mathcal{L}}))$ where $\tilde{L} \equiv \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$ and*

$$\tilde{f}(\cdot; \underline{\mathcal{L}}) : \mathcal{P}(\tilde{L}) \to \mathbb{R} \qquad , \qquad \tilde{f}(\mathcal{L}; \underline{\mathcal{L}}) \equiv f(\mathcal{L} \cup \underline{\mathcal{L}}).$$

Proof. We show that $\mathcal{L}' \setminus \underline{\mathcal{L}} = \Phi(\mathcal{L}' \setminus \underline{\mathcal{L}}; \tilde{f}(\cdot; \underline{\mathcal{L}}))$. Let $\ell \in \mathcal{L}' \setminus \underline{\mathcal{L}}$. Then, $\ell \notin \underline{\mathcal{L}}$ and also $\ell \in \mathcal{L}' = \Phi(\mathcal{L}'; f)$ implies that $D_\ell f(\mathcal{L}') = D_\ell \tilde{f}(\mathcal{L}' \setminus \underline{\mathcal{L}}; \underline{\mathcal{L}}) \geq 0$ so $\ell \in \Phi(\mathcal{L}' \setminus \underline{\mathcal{L}}; \tilde{f}(\cdot; \underline{\mathcal{L}}))$. Now let $\ell \in \Phi(\mathcal{L}' \setminus \underline{\mathcal{L}}; \tilde{f}(\cdot; \underline{\mathcal{L}}))$. Then, $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$ and $D_\ell \tilde{f}(\mathcal{L}' \setminus \underline{\mathcal{L}}; \underline{\mathcal{L}}) = D_\ell f(\mathcal{L}') \geq 0$ so $\ell \in \Phi(\mathcal{L}'; f) = \mathcal{L}'$ and $\ell \notin \underline{\mathcal{L}}$ so $\ell \in \mathcal{L}' \setminus \underline{\mathcal{L}}$. □

Theorem. *Let $\Xi \equiv \{\mathcal{L} \mid \mathcal{L} = \Phi(\mathcal{L})\}$ be the set of fixed points of Φ. Then, $\Xi \subseteq \Lambda$ for any outcome of the branching procedure Λ.*

Proof. Let $\mathcal{L}' \in \Xi$ be an arbitrary fixed point of Φ. The reduced bounding sets are the smallest and largest fixed points of Φ when f satisfies SCD-C from below, and the extreme edge of Φ when f satisfies SCD-C from above. Thus, $\mathcal{L}'$ is contained in the reduced domain. We use strong induction on the cardinality of $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$.

Suppose $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}} = 0$. Then, $\underline{\mathcal{L}} = \mathcal{L}^\star = \overline{\mathcal{L}}$ is the unique fixed point of Φ and the statement holds trivially.

Now, suppose the statement holds if there are k elements in $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$. Consider a scenario where $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$ contains $k+1$ elements. Branching partitions the remaining decision sets into two disjoint collections, the one defined by $[\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}]$ and the one defined by $[\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}]$.

Exactly one of the two branches accords with $\mathcal{L}$, depending on whether or not it contains ℓ. Then, consider the relevant branch. By the previous lemma, $(\mathcal{L}' \setminus \underline{\mathcal{L}}) \setminus$

$\{\ell\}$ is a fixed point of the branch's conditional objective function. Thus, applying the squeezing procedure on this branch does not squeeze out $(\mathcal{L}' \setminus \underline{\mathcal{L}}) \setminus \{\ell\}$.

If the squeezing procedure applied to this branch converges to $(\mathcal{L}' \setminus \underline{\mathcal{L}}) \setminus \{\ell\}$, then this branch terminates with $\mathcal{L}'$ so $\mathcal{L}' \in \Lambda$. If the squeezing procedure does not converge to a single decision set, the reduced domain on this branch is such that the number of items that separate the reduced bounding sets cannot exceed k. Then, applying the inductive assumption, $\mathcal{L}'$ is on one of the terminal nodes of the branching tree as branching continues. □

Corollary. *The branching procedure, regardless of items selected for branching, identifies a global optimum* $\mathcal{L}^\star \in \arg\max_{\mathcal{L} \in \mathcal{P}(L)} f(\mathcal{L})$.

Proof. Let $\Xi \equiv \{\mathcal{L} \mid \mathcal{L} = \Phi(\mathcal{L})\}$ be the set of fixed points of Φ and Λ be an outcome of the branching procedure. From the previous theorem, $\Xi \subseteq \Lambda$. From a previous lemma, the set of fixed points Ξ contains a global maximizer, so that

$$\max_{\mathcal{L} \in \Xi} f(\mathcal{L}) = \max_{\mathcal{L} \in \mathcal{P}(L)} f(\mathcal{L}).$$

Then, the branching procedure's outcome, Λ, must contain a global maximizer. Since Λ was an arbitrary outcome of the branching procedure, the branching procedure must identify a global maximizer regardless of the items selected for branching. □

E.3. Proofs: Heterogeneous Agents

Policy Function

Definition (Finite crossing differences in type (FCD-T))**.** For any pair of strategies $(\mathcal{L}, \mathcal{L}')$, the set $\{z \mid f(\mathcal{L}, z) - f(\mathcal{L}', z) = 0\}$ is finite.

The FCD-T condition is a generalization of strong SCD-T.

Theorem (Continuous policy function almost everywhere)**.** *Suppose FCD-T holds and L is finite. Then,*

1. *The set of types where* $\mathcal{L}^*(\cdot)$ *is not unique is finite, so that* $\mathcal{L}^\star(\cdot)$ *is a function almost everywhere.*
2. *If* $\mathcal{L}^\star(z)$ *is unique, there exists* $\delta > 0$ *so that for all* $z' \in [z - \delta, z + \delta]$, *the optimal strategy is also* $\mathcal{L}^\star(z)$, *so that* $\mathcal{L}^\star(\cdot)$ *is continuous at these points.*

3. *If the optimal strategy for z is not unique,*

   a) *there exists a unique strategy $\mathcal{L}_-$ where there exists a $\delta_- > 0$ so that $\mathcal{L}_-$ is optimal for all $z' \in [z-\delta, z]$, so that $\mathcal{L}^\star(\cdot)$ is left-continuous; and*

   b) *there exists a unique strategy $\mathcal{L}_+$ where there exists a $\delta_+ > 0$ so that $\mathcal{L}_+$ is optimal for all $z' \in [z, z+\delta]$, so that $\mathcal{L}^\star(\cdot)$ is right-continuous.*

Proof. We show each statement in turn.

1. If there are zero or one types where $\mathcal{L}^\star(z)$ is not unique, then the theorem holds trivially. Suppose there is more than one type where the optimal strategy is not unique. Let

$$n = \max_{(\mathcal{L},\mathcal{L}')} \left|\left\{z \mid f(\mathcal{L}, z) = f\left(\mathcal{L}', z\right)\right\}\right|$$

   where note the maximum exists because there are a finite number of pairs $(\mathcal{L}, \mathcal{L}')$. Then, there cannot be more than $n \times |\mathcal{P}(\mathcal{P}(L))|$ additional types where the optimal strategy is not unique.

2. Define

$$g\left(z'\right) = \min_{\mathcal{L} \neq \mathcal{L}^\star(z)} \left\{f\left(\mathcal{L}^\star(z), z'\right) - f\left(\mathcal{L}, z'\right)\right\}.$$

   At z, $g(z) > 0$ since $\mathcal{L}^\star(z)$ is unique. It is continuous since differences of continuous functions are continuous, and so g is a minimum over a set of continuous function. Then, there is a $\tilde{\delta} > 0$ so that, for all $z' \in \left(z-\tilde{\delta}, z+\tilde{\delta}\right)$,

$$\begin{aligned}
&\left|g\left(z'\right) - g(z)\right| < \frac{1}{2} g(z) \\
&\qquad \Rightarrow g\left(z'\right) > g(z) - \frac{1}{2} g(z) = \frac{1}{2} g(z) > 0\,.
\end{aligned}$$

   Then, for all $z' \in \left(z-\tilde{\delta}, z+\tilde{\delta}\right)$,

$$0 < f\left(\mathcal{L}^\star(z), z'\right) - f\left(\mathcal{L}, z'\right) \qquad\qquad \forall \mathcal{L} \neq \mathcal{L}^\star(z)$$

   so $\mathcal{L}^\star(z)$ is uniquely optimal on this interval. Then, setting $\delta = \frac{1}{2}\tilde{\delta}$ is sufficient.

3. We prove the right-continuity statement; the proof of the left-continuity statement is identical in spirit. Use induction on the number of optimal strategies. Suppose there are exactly two optimal strategies $\mathcal{L}_1$ and $\mathcal{L}_2$ for the type z. For $i \in \{1, 2\}$, let

$$g_i\left(z'\right) = \min_{\mathcal{L} \notin \{\mathcal{L}_1, \mathcal{L}_2\}} \left\{f\left(\mathcal{L}_i, z'\right) - f\left(\mathcal{L}, z'\right)\right\}.$$

By the same arguments as above, the function $\tilde{g} \equiv g_2 - g_1$ is continuous. First, suppose $\tilde{g}(z') \neq 0$ for all $z' > z$. By the intermediate value theorem, it must either be that $\tilde{g}(z') > 0$ or $\tilde{g}(z') < 0$. Assign

$$\mathcal{L}_+ = \begin{cases} \mathcal{L}_2 & \text{if } \tilde{g}(z') > 0 \text{ for } z' > z \\ \mathcal{L}_1 & \text{if } \tilde{g}(z') < 0 \text{ for } z' > z \end{cases}$$

Note that $\mathcal{L}_+$ is uniquely optimal for all $z' > z$. Then, the statement holds for any positive value of δ. Now, suppose $\tilde{g}(z') = 0$ for some $z' > z$. Let

$$\underline{z} \equiv \min\left\{z' \mid z' > z, \tilde{g}\left(z'\right) = 0\right\} > z$$

which exists since the set is non-empty and finite by finite crossing. Set $\tilde{\delta} = \frac{1}{2}(\underline{z} - z)$. Then, for any $z' \in \left(z, z + \tilde{\delta}\right)$, we have that $\tilde{g}(z') \neq 0$. In the same argument as above, $\tilde{g}(z') > 0$ for the entire interval or $\tilde{g}(z') < 0$. Thus, assigning $\delta = \frac{1}{4}(\underline{z} - z)$ is sufficient. We have now established the base case.

Now suppose the statement holds if there are k optimal strategies at z. Consider the case where there are $k+1$ optimal strategies, enumerated $\mathcal{L}_1, \ldots, \mathcal{L}_{k+1}$, and define the modified objective function

$$\tilde{f}(\mathcal{L}, z') = \begin{cases} f(\mathcal{L}, z') & \text{if } \mathcal{L} \neq \mathcal{L}_{k+1} \\ f(\mathcal{L}_1, z') - 1 & \text{if } \mathcal{L} = \mathcal{L}_{k+1} \end{cases}.$$

At z, the first k strategies are optimal for the modified problem, but $\mathcal{L}_{k+1}$ is not. Then, the statement holds for the modified problem; let $\mathcal{L}_+ \in \{\mathcal{L}_i \mid i = 1, \ldots, k\}$ be the strategy for which the statement holds in the modified problem. Repeat the proof for the base case with the objective f, assigning $\mathcal{L}_1 = \mathcal{L}_+$ and $\mathcal{L}_2 = \mathcal{L}_{k+1}$.

□

Single Crossing Differences in Type

Proposition. *Consider the objective function $f : \mathcal{P}(L) \times \mathbb{R} \to \mathbb{R}$. The single crossing property, as defined in Milgrom and Shannon [1994], is sufficient for SCD-T.*

Proof. Suppose f has the single crossing property, that is, for any $\mathcal{L} \subset \mathcal{L}'$ and $z < z'$,

$$f\left(\mathcal{L}', z\right) - f\left(\mathcal{L}, z\right) \geq 0 \qquad \Rightarrow \qquad f\left(\mathcal{L}', z'\right) - f\left(\mathcal{L}, z'\right) \geq 0$$

and choose an arbitrary $\mathcal{L} \in \mathcal{P}(L)$, $\ell \in L$. Let $z < z'$ and suppose $D_\ell f(\mathcal{L}, z) \geq 0$. Then,

$$\begin{aligned} D_\ell f(\mathcal{L}, z) &= f(\mathcal{L} \cup \{\ell\}, z) - f(\mathcal{L} \setminus \{\ell\}, z) \geq 0 \\ \Rightarrow D_\ell f(\mathcal{L}, z') &= f(\mathcal{L} \cup \{\ell\}, z') - f(\mathcal{L} \setminus \{\ell\}, z') \geq 0 \end{aligned}$$

so f satisfies SCD-T. □

The Generalized Squeezing Procedure

Lemma. *Suppose the underlying objective function f satisfies SCD-T. Let $\ell \in L$ and $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ where $\mathcal{L} \subset \mathcal{L}'$.*

1. *If f also satisfies SCD-C from above, then $z_\ell^g(\mathcal{L}) \leq z_\ell^g(\mathcal{L}')$.*
2. *If f also satisfies SCD-C from below, then $z_\ell^g(\mathcal{L}') \leq z_\ell^g(\mathcal{L})$.*

Proof. If f satisfies SCD-C from above,

$$0 = D_\ell\left(\mathcal{L}', z_\ell^g(\mathcal{L}')\right) \overset{\text{SCD-C from above}}{\Rightarrow} 0 \leq D_\ell\left(\mathcal{L}, z_\ell^g(\mathcal{L}')\right) \overset{\text{SCD-T}}{\Rightarrow} z_\ell^g(\mathcal{L}) \leq z_\ell^g(\mathcal{L}');$$

if f satisfies SCD-C from below,

$$0 = D_\ell\left(\mathcal{L}, z_\ell^g(\mathcal{L})\right) \overset{\text{SCD-C from below}}{\Rightarrow} 0 \leq D_\ell\left(\mathcal{L}', z_\ell^g(\mathcal{L})\right) \overset{\text{SCD-T}}{\Rightarrow} z_\ell^g(\mathcal{L}') \leq z_\ell^g(\mathcal{L}).$$

□

From the previous lemma, if f satisfies SCD-T, the generalized squeezing step simplifies to

$$S^g\left(\left[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)\right]\right) \equiv \begin{cases} \left[\Phi^g\left(\overline{\mathcal{L}}(\cdot), \cdot\right), \Phi^g\left(\underline{\mathcal{L}}(\cdot), \cdot\right)\right] & \text{with SCD-C from above} \\ \left[\Phi^g\left(\underline{\mathcal{L}}(\cdot), \cdot\right), \Phi^g\left(\overline{\mathcal{L}}(\cdot), \cdot\right)\right] & \text{with SCD-C from below} \end{cases}.$$

We now prove Theorem 2.

Proof. We prove the case of SCD-C from above. The argument in the case of SCD-C from below follows the same logic.

Let $\Phi(\mathcal{L}, z) \equiv \{\ell \mid D_\ell f(\mathcal{L}, z) \geq 0\}$ be the mapping Φ evaluated at the type z. Consider a pair of bounding functions $\left[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)\right]$ and let the bounding sets at this type be $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right] \equiv \left[\underline{\mathcal{L}}(z), \overline{\mathcal{L}}(z)\right]$.

Applying Theorem 1 at z, we have $\underline{\mathcal{L}} \subseteq \Phi\left(\overline{\mathcal{L}}, z\right) \subseteq \mathcal{L}^{\star}(z) \subseteq \Phi\left(\underline{\mathcal{L}}, z\right) \subseteq \overline{\mathcal{L}}$. Thus, it is sufficient to show that $\Phi^{g}\left(\overline{\mathcal{L}}, z\right) = \Phi\left(\overline{\mathcal{L}}, z\right)$ and $\Phi^{g}\left(\underline{\mathcal{L}}, z\right) = \Phi\left(\underline{\mathcal{L}}, z\right)$. Let $\ell \in \Phi^{g}\left(\overline{\mathcal{L}}, z\right)$. Then, $z_{\ell}^{g}\left(\overline{\mathcal{L}}\right) \leq z$ and $0 = D_{\ell} f\left(\overline{\mathcal{L}}, z_{\ell}^{g}\left(\overline{\mathcal{L}}\right)\right)$ together imply $0 \leq D_{\ell}\left(\overline{\mathcal{L}}, z\right)$ by SCD-T. Thus, $\ell \in \Phi\left(\overline{\mathcal{L}}, z\right)$. Similarly, if $\ell \in \Phi\left(\overline{\mathcal{L}}, z\right)$, then $0 \leq D_{\ell} f\left(\overline{\mathcal{L}}, z\right)$ implies $z \geq z_{\ell}^{g}\left(\overline{\mathcal{L}}\right)$ by SCD-T. Thus, $\ell \in \Phi^{g}\left(\overline{\mathcal{L}}, z\right)$. Thus, $\Phi^{g}\left(\overline{\mathcal{L}}, z\right) = \Phi\left(\overline{\mathcal{L}}, z\right)$; a similar argument establishes $\Phi^{g}\left(\underline{\mathcal{L}}, z\right) = \Phi\left(\underline{\mathcal{L}}, z\right)$.

Similarly, since each application of the generalized squeezing step is equivalent to applying the squeezing step point-wise, the generalized squeezing procedure converges in at most $|L|$ iterations. □

Corollary. *When f satisfies SCD-C and SCD-T, the generalized squeezing procedure converges to bounding functions $\left[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)\right]$ where, for every type z:*

1. *$\underline{\mathcal{L}}(z)$ and $\overline{\mathcal{L}}(z)$ are the smallest and largest fixed points of $\Phi(\cdot, z)$ if f satisfies SCD-C from below; and*
2. *$\underline{\mathcal{L}}(z)$ and $\overline{\mathcal{L}}(z)$ is the extreme fixed edge of $\Phi(\cdot, z)$ if f satisfies SCD-C from above.*

Proof. Follows from $\Phi^{g}(\mathcal{L}, z) = \Phi(\mathcal{L}, z)$ for all $\mathcal{L}$ and z and previous results. □

Corollary. *Suppose f satisfies SCD-C and SCD-T, and let $\left[\mathcal{L}^{\inf}(z), \mathcal{L}^{\sup}(z)\right]$ be the smallest and largest fixed points of $\Phi(\cdot, z)$ (with SCD-C from below) or the extreme fixed edge (with SCD-C from above). The functions $\left[\underline{\mathcal{L}}^{\inf}(\cdot), \overline{\mathcal{L}}^{\sup}(\cdot)\right]$ change value a finite number of times.*

Proof. By the previous corollary, the generalized squeezing procedure converges to $\left[\mathcal{L}^{\inf}(\cdot), \mathcal{L}^{\sup}(\cdot)\right]$. We thus prove the statement for the generalized squeezing procedure's output, using induction on the iteration number k.

At $k = 0$, the domain bounding functions $\left[\underline{\mathcal{L}}^{(0)}(\cdot), \overline{\mathcal{L}}^{(0)}(\cdot)\right] = [\emptyset, L]$ never change value.

Suppose the bounding functions $\left[\underline{\mathcal{L}}^{(k)}(\cdot), \overline{\mathcal{L}}^{(k)}(\cdot)\right]$ at iteration k each only change value a finite number of time. Then, they induce a finite partition of intervals on the type space

$$\mathcal{T}\left(\left[\underline{\mathcal{L}}^{(k)}(\cdot), \overline{\mathcal{L}}^{(k)}(\cdot)\right]\right) = \{\mathcal{Z}_1, \ldots \mathcal{Z}_t, \ldots \mathcal{Z}_T\}$$
$$\text{such that } \mathcal{Z}_t = \left\{z \in \mathbb{R} \mid \underline{\mathcal{L}}^{(k)}(z) = \underline{\mathcal{L}}_t, \overline{\mathcal{L}}^{(k)}(z) = \overline{\mathcal{L}}_t\right\},$$

where the functions are constant over each interval. Consider one such interval $\mathcal{Z}_t$. Since $\underline{\mathcal{L}}_t \subseteq \underline{\mathcal{L}}^{(k+1)}(z) \subseteq \overline{\mathcal{L}}^{(k+1)}(z) \subseteq \overline{\mathcal{L}}_t$ for all $z \in \mathcal{Z}_t$,

$$z_\ell^g\left(\underline{\mathcal{L}}_t\right) \notin \mathcal{Z}_t \qquad \text{and} \qquad z_\ell^g\left(\overline{\mathcal{L}}_t\right) \notin \mathcal{Z}_t \qquad \forall \ell \notin \overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t$$

so $(k+1)$th application of the generalized squeezing step adds at most $2 \times \left|\overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t\right|$ points within the interval where one of the bounding functions change value. Overall, the new bounding functions $\left[\underline{\mathcal{L}}^{(k+1)}(\cdot), \overline{\mathcal{L}}^{(k+1)}(\cdot)\right] = S^g\left(\left[\underline{\mathcal{L}}^{(k)}(\cdot), \overline{\mathcal{L}}^{(k)}(\cdot)\right]\right)$ change value at most

$$2\sum_{t=1}^{T}\left|\overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t\right|$$

more times than the bounding functions at iteration k. □

Policy Function Refinement

Corollary. *If f satisfies SCD-C and SCD-T, the generalized squeezing branching procedure converges to*

$$\Xi(z) = \{\mathcal{L} \mid \mathcal{L} = \Phi(\mathcal{L}, z)\},$$

the set of all fixed points of $\Phi(\mathcal{L}, z)$ at z.

Proof. Follows from the fact that, with SCD-C and SCD-T, the generalized branching step coincides pointwise with the branching step. □

Lemma. *Suppose f satisfies strong SCD-T. Then,*

1. *For any z, z' in the type space, if there is a set $\mathcal{L}$ that is optimal (not necessarily uniquely) at both z and z', then it is optimal for types on the interval $[z, z')$.*
2. *At all iterations k of iterative cutoff search: the end points satisfy $z^{(0)} \leq z^{(k)} < z'^{(k)} \leq z'^{(0)}$; and the indifferent type $\overline{z}^{(k)}$ exists, is unique, and is within $\left[z^{(k)}, z'^{(k)}\right]$.*

Proof. We prove each statement in turn.

1. Suppose $\mathcal{L}$ is optimal at both z and z'. For a contradiction, suppose there exists a type $\overline{z} \in (z, z')$ for which $\mathcal{L}$ is not optimal. Let $\mathcal{L}'$ be optimal at $\overline{z}$. Define $g(z) \equiv f(\mathcal{L}', z) - f(\mathcal{L}, z)$. Then, g is continuous and by assumption $g(\overline{z}) > 0$. In addition, $g(z) \leq 0$ and $g(z') \leq 0$ by construction. Then, by the intermediate value theorem, g crosses 0 at least once on the interval $[z, \overline{z})$ and again on $(\overline{z}, z']$, violating strong SCD-T.

2. We start with the indifferent type statement. At iteration k, define

$$g(z) = f\left(\mathcal{L}^{\star}\left(z^{(k)}\right), z\right) - f\left(\mathcal{L}^{\star}\left(z'^{(k)}\right), z\right).$$

By strong SCD-T, there is at most one type $\overline{z}^{(k)}$ where $g\left(\overline{z}^{(k)}\right) = 0$; but also, since $g\left(z^{(k)}\right) \leq 0$ and $g\left(z'^{(k)}\right) \geq 0$, the intermediate value theorem guarantees that this type exists and is in between $z^{(k)}$ and $z'^{(k)}$.

We now establish the end point statement, using induction on k. At $k = 0$, $z^{(0)} < z'^{(0)}$ trivially. Suppose the statement is true for iteration k. Consider iteration $(k+1)$. Then, either

$$z^{(k+1)} = z'^{(k)} \neq z'^{(0)} \quad , \quad z'^{(k+1)} = z'^{(0)}$$

or

$$z^{(k+1)} = z^{(k)} \quad , \quad z'^{(k+1)} = \overline{z}^{(k)}$$

from the previous iteration. If the first is true, by inductive assumption, $z^{(0)} \leq z'^{(k)} < z'^{(0)}$ so $z^{(0)} \leq z^{(k+1)} < z'^{(k+1)} = z'^{(0)}$. If the second is true, by inductive assumption and the indifferent type statement, $z^{(0)} \leq z^{(k)} = z^{(k+1)} \leq \overline{z}^{(k)} = z'^{(k+1)} \leq z'^{(0)}$.

□

Theorem. *If f satisfies strong SCD-T, then iterative cutoff search:*

1. *correctly identifies $\mathcal{L}^{\star}(\cdot)$ on $\left[z^{(0)}, z'^{(0)}\right)$; and*
2. *completes in at most $\binom{\left|\overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t\right|}{2}$ iterations.*

Proof. Note that iterative search concludes in a finite number of iterations, since the reduced domain is finite and strong SCD-T guarantees at most $\binom{\left|\overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t\right|}{2}$ cutoffs, proving the second statement.

To prove the first statement, use strong induction on the iteration at which iterative cutoff search concludes. Consider the case where the iteration terminates at the $k = 0$ iteration. By the previous lemma, $\overline{z}^{(0)} \in \left[z^{(0)}, z'^{(0)}\right]$. Since iteration terminates, $\mathcal{L}^{\star}\left(\overline{z}^{(0)}\right)$ must coincide with either $\mathcal{L}^{\star}\left(z^{(0)}\right)$ or $\mathcal{L}^{\star}\left(z'^{(0)}\right)$. Then,

$$f\left(\mathcal{L}^{\star}\left(\overline{z}^{(0)}\right), \overline{z}^{(0)}\right) = f\left(\mathcal{L}^{\star}\left(z^{(0)}\right), \overline{z}^{(0)}\right) = f\left(\mathcal{L}^{\star}\left(z'^{(0)}\right), \overline{z}^{(0)}\right).$$

By the previous lemma, the policy function is set correctly along the entire interval.

Suppose, if the search concludes at iteration 0, 1, up to n, then it correctly identifies the policy function. Consider the case where it concludes at the iteration $(n+1)$. Then, when $k = 0$, the optimal set at $\overline{z}^{(0)}$ cannot coincide with either end point optimal set, since iteration would terminate. Then, it must be that

$$z^{(1)} = z^{(0)} \qquad , \qquad z'^{(1)} = \overline{z}^{(0)}.$$

Since the full iteration concludes when $k = (n+1)$, there must be some $k' \leq n$ where $z^{(k'+1)} = \overline{z}^{(0)}$. Consider alternatively initiating iterative cutoff search on $\left[z^{(0)}, \overline{z}^{(0)}\right)$ and call this search "left search." Note that left search proceeds identically to the original iterations $k \in \{1, \ldots, k'\}$, so left search terminates by at least iteration n. By strong inductive assumption, left search correctly determines the policy function on $\left[z^{(0)}, \overline{z}^{(0)}\right)$. Then, iterations $k \in \{1, \ldots, k'\}$ of the original iterative cutoff search correctly identify the policy function on $\left[z^{(0)}, \overline{z}^{(0)}\right)$.

Once left search has concluded, iterative cutoff search sets $z^{(k'+1)} = \overline{z}^{(0)}$ and $z'^{(k'+1)} = z'^{(0)}$. Iterations $(k'+1)$ to $(n+1)$ proceed as if the iteration had been initiated with $\left[\overline{z}^{(0)}, z'^{(0)}\right)$. By a similar argument above, the strong inductive assumption guarantees that iterative cutoff search correctly determines the policy function on this interval. □

F. Model Details

In this section, we lay out a microfoundation for the CES marginal cost function in equation 1 and relate the cross-location employment elasticity to SCD-C.

F.1. Input Microfoundation for CES Marginal Cost

Firms produce their final good by combining a continuum of firm-specific intermediate inputs, indexed by m, with a constant elasticity of substitution η. Each of the firm's production locations can produce the entire continuum of intermediate inputs.

For a firm headquartered in location i, the marginal cost of producing an input variety m at a production site in location ℓ is given by $\gamma_{i\ell} w_\ell / \varphi_\ell(m)$, where $\varphi_\ell(m)$ is a location-input-specific productivity shock and $\gamma_{i\ell}$ is a bilateral cost of multinational production. For each destination n and intermediate input m, the firm chooses from among its set of production sites, $\mathcal{L}$, the location $\ell_{in}^{\star}(\boldsymbol{\varphi}(m))$ that offers the lowest destination-specific marginal cost:

$$\ell_{in}^{\star}(\boldsymbol{\varphi}(m)) = \arg\min_{\ell \in \mathcal{L}} \gamma_{i\ell} \frac{w_\ell}{\varphi_\ell(m)} \tau_{\ell n},$$

where the term $\tau_{\ell n}$ denotes a bilateral iceberg trade costs and the vector $\boldsymbol{\varphi}(m) = \{\varphi_\ell(m)\}_\ell$ collects the firm's productivities of producing the input m in each location ℓ.

Suppose the firm draws the productivity terms $\varphi_\ell(m)$ independently from a Fréchet distribution with shape θ and scale $A_\ell z$, *after* making its production location decision $\mathcal{L}$. Then, the Fréchet distribution on idiosyncratic location draws implies the CES cost function used in the main body of the paper with $\varepsilon = 1 + \theta$, up to a constant of integration. The substitutability among plants derives from the fact that plants cannibalize one another's sales as they compete to be the least cost supplier. The strength of this force depends on how much production locations differ in their productivity at producing any given variety as measured by the dispersion $(1/\theta)$ of the location-input-specific productivity shocks. If comparative advantage differences among production locations are large ($1/\theta$ is large), the substitutability across locations is low and cannibalization is limited.

The properties of the Fréchet distribution imply that the expression in equation (1) is independent of the elasticity of aggregation across varieties η [see Eaton and Kortum, 2002]. A similarly tractable expression arises if, instead of the independent Fréchet distributions, each location-input-specific productivity shock is drawn from either a multivariate correlated Fréchet [as in Ramondo, 2014] or a multivariate correlated Pareto distribution [as in Arkolakis et al., 2018]. Integrating across inputs delivers the CES cost function in (1), with θ replaced by $\frac{\theta}{1-\rho}$, where θ is the shape and ρ is the correlation.

F.2. SCD-C and Cross-Location Employment Elasticity

In Section 2, we provide the parameter restriction which determines whether locations are complements or substitutes in the model. In particular, if the elasticity of substitution among locations in the firm's cost function, ε, exceeds the elasticity of demand, σ, then locations are substitutes; otherwise they are complements.

We give an employment elasticity interpretation of this restriction. Consider the partial equilibrium response of the firm's variable production employment at location $\ell' \in \mathcal{L}$ to a small change in the wage w_ℓ in location $\ell \in \mathcal{L}$, holding fixed the firm's decision set $\mathcal{L}$ and all other aggregates. This employment elasticity is

$$\left.\frac{\partial \ln\left(\text{emp. in } \ell'\right)}{\partial \ln w_\ell}\right|_{\mathcal{L}} = (\sigma - \varepsilon)\sum_n \left[\frac{s_{i\ell' n}\left(\mathcal{L}\right) y_{in}\left(\mathcal{L}, z\right)}{\sum_{n'} s_{i\ell' n'}\left(\mathcal{L}\right) y_{in'}\left(\mathcal{L}, z\right)}\right] s_{i\ell n}\left(\mathcal{L}\right) - (\varepsilon - 1)\,\mathbb{1}\left[\ell = \ell'\right]$$

where $y_{in}\left(\mathcal{L}, z\right) = q_n\left(p_{in}\left(\mathcal{L}, z\right)\right) p_{in}\left(\mathcal{L}, z\right)$ are the firm's total sales in market n and $s_{i\ell n}\left(\mathcal{L}\right)$ is the share of sales to n that is produced in ℓ.

The own-elasticity, when $\ell' = \ell$, is always negative: as the wage increases, the firm adjusts downwards its employment in that location, all else equal. However, the sign of the cross-location employment elasticity, when $\ell' \neq \ell$, depends on the size of ε relative to σ. If the cross-elasticity is positive, so that the firm increases employment at all other locations when the wage at a given location increases, then locations act as substitutes. This case corresponds to $\varepsilon > \sigma$, the condition for the firm's problem to satisfy SCD-C from above. On the other hand, if the firm instead decreases employment at all other locations, then locations are complements. This case corresponds to $\varepsilon < \sigma$, when the firm's problem satisfies SCD-C from below.

These elasticities are the related to those estimated in Muendler and Becker [2010], but not directly comparable since our elasticities depend on the firm's particular location set $\mathcal{L}$ and the location-market shares of each (ℓ, n) pair.

G. Generalized Theoretical Framework

In this section, we relax the assumptions on the production structure and the demand system in our model in Section 2. The generalized cost and demand functions nest several frameworks. We then show how to establish both single crossing differences conditions in this more general setup.

G.1. General Cost Function

Consider a firm of productivity $z \in \mathbb{R}^+$ headquartered in country i with a production location set $\mathcal{L}$. In what follows, we omit the i index for brevity.

Let $c_n\left(\mathcal{L}, z\right)$ be the unit cost of delivering its final good to a destination market n. Section 2 presents a particular formulation for $c_n\left(\mathcal{L}, z\right)$, microfounded by Fréchet location-input-specific cost shocks. Here, we relax the assumption on $c_n\left(\mathcal{L}, z\right)$ while remaining agnostic on its microfoundation.

Assumption 1 (Generalized marginal cost function)**.** The marginal cost function of a firm with productivity z in destination n can be written as the composition $c_n(\mathcal{L}) = g(\boldsymbol{\Theta}_n(\mathcal{L}), z)$ of the vector-valued production index function $\boldsymbol{\Theta}_n : \mathcal{P}(\mathcal{L}) \to \mathbb{R}^K$ and the outer cost function $g : \mathbb{R}^K \times \mathbb{R}^+ \to \mathbb{R}^+$ where:

1. each dimension of the production index function features no interdependencies among elements of $\mathcal{L}$ and is increasing in $\mathcal{L}$, i.e. for all $\ell \in L$, $\mathcal{L} \subseteq L$, and $k \leq K$,

$$\tilde{\xi}_{k\ell n} \equiv \Theta_{kn}(\mathcal{L} \cup \{\ell\}) - \Theta_{kn}(\mathcal{L} \setminus \{\ell\}) \geq 0$$

   is independent of $\mathcal{L}$, where $\Theta_{kn}(\mathcal{L})$ is the kth entry of $\boldsymbol{\Theta}_n(\mathcal{L})$;

2. for every production location ℓ, there is at least one market n and dimension of production potential k that is strictly improved by ℓ, i.e. $\tilde{\xi}_{k\ell n} > 0$; and

3. the outer cost function g is twice-differentiable and monotonically decreasing in each dimension of production potential and in firm productivity (if it is increasing, redefine $\tilde{z} \equiv -z$), i.e. for all $\mathcal{L} \subseteq L$, $k \leq K$, and $z \in \mathbb{R}^+$

$$\frac{\partial g}{\partial \Theta_{kn}} < 0 \qquad \text{and} \qquad \frac{\partial g}{\partial z} < 0\,.$$

The central object in Assumption 1 is the production "index" function $\boldsymbol{\Theta}_n$ that measures the overall potential of the production location set $\mathcal{L}$ along K dimensions. These dimensions could represent different technological techniques of production, industries, or other latent variables. Previous work in the multinational literature refers to $\boldsymbol{\Theta}_n(\mathcal{L})$ as the "production potential" or "sourcing potential" associated with a given location set [see, e.g., Antràs et al., 2017]. For each dimension k, the marginal contribution of each location to the index is independent of the marginal contribution of other locations.

Assumption 2 (General fixed cost function)**.** The total fixed cost of establishing a production location set $\mathcal{L}$ for a firm headquartered in location i, $F(\mathcal{L})$, is given by:

$$F(\mathcal{L}) = \sum_{\ell \in \mathcal{L}} F_\ell.$$

This assumption asserts that there is an independent fixed cost F_ℓ for establishing each production location. In the framework of Section 2, these fixed costs are $F_{i\ell} = w_\ell f_{i\ell}$ for firms headquartered in i.

Special Cases We discuss existing frameworks which satisfy the structure imposed in Assumption 1. Consider first the case where $K = 1$, so the production index $\Theta_n(\mathcal{L})$ is a scalar. The CES marginal cost function $c_n(\mathcal{L}, z)$ in Section 2, as well as in Tintelnot [2017], Antràs et al. [2017], follow this structure with

$$\Theta_n(\mathcal{L}) = \sum_{\ell \in \mathcal{L}} \xi_{\ell n}^{1-\varepsilon} \quad , \quad g(\Theta, z) = \frac{\Gamma}{z} \Theta^{\frac{1}{1-\varepsilon}}$$

where $\tilde{\xi}_{\ell n} = \xi_{\ell n}^{1-\varepsilon}$ is a combination of fundamentals and aggregates, Γ is a constant of integration, and $\varepsilon = 1 + \theta$. As discussed in the location-input-specific cost shock microfoundation, Assumption 1 is also satisfied in models in which the location-input-specific productivity shocks are distributed according to a multivariate Pareto as in Arkolakis et al. [2018], or a Fréchet distribution with a uniform correlation across draws. The elasticity of substitution is $\varepsilon = 1 + \frac{\theta}{1-\rho}$.

Lind and Ramondo [2023] present a cost function that features K nests. This cost function satisfies the multidimensional case with $K > 1$. In this case,

$$\Theta_{kn}(\mathcal{L}) = \sum_{\mathcal{L}} \omega_{kn}^{\frac{1}{1-\rho_k}} \xi_{k\ell n}^{-\frac{\theta}{1-\rho_k}} \quad , \quad g(\boldsymbol{\Theta}, z) = \frac{\Gamma}{z} \left[\sum_k \Theta_{kn}(\mathcal{L})^{1-\rho_k} \right]^{-\frac{1}{\theta}}$$

where now the location set $\mathcal{L}$ maps to a different potential for each technique k, the weights ω_{kn} describe the importance of each technique k to destination n, and ρ_k is the substitutability across locations within the nest k. As an example, suppose there is a standard production technique $k = 1$ and a skill-intensive production technique $k = 2$. In this case, $K = 2$, and the nests represent production techniques. Then, $\Theta_{1n}(\mathcal{L})$ represents overall potential of the location set $\mathcal{L}$ when the firm applies the standard production technique, while $\Theta_{2n}(\mathcal{L})$ represents the potential when applying the skill-intensive technique. These potentials differ since the low-skill and high-skill wages could differ in each location, so the potential of a particular location set depends on which technique the firm uses. In the same way, locations substitute for each other within each production technique, captured by ρ_k, but not directly across techniques. Lind and Ramondo [2023] microfound this cost function using a nested multivariate Fréchet distribution with shape θ and correlations ρ_k, then show that it is sufficiently flexible to approximate any general correlation structure up to arbitrarily close precision. This formulation nests the previous special cases when $K = 1$.

G.2. General Demand Function

Consider a set of destination markets N, each of which feature consumers with residual demand function $q_n(p_n)$. We then impose the following structure on the firm's variable profits.

Assumption 3 (Generalized variable profit function)**.** Conditional on market aggregates, the variable profits of a firm headquartered in location i take the form

$$v\left(\mathbf{c};\{p_n\}_n\right)=\sum_{n\in N}\left[q_n\left(p_n\right)p_n-q_n\left(p_n\right)c_n\right],$$

where $\mathbf{c}=\{c_n\}_n$ is the vector of unit costs of producing and delivering a good to the destination markets n, p_n is the price charged in market n, and the residual demand function q_n is differentiable and monotonically decreasing.

A key feature of this profit function is that the destination markets are independent and that there are no strategic interactions among firms. In particular, the unit cost c_n of serving a market n does not affect the variable profits earned in destination market $n'\neq n$. Similarly, the price p_n set by the firm in market n does not affect the variable profits in a different destination market. This formulation does not require demand to be homothetic, nor does it place any particular restrictions on the elasticity of demand.

Following standard firm maximization, the firm sets a different price in each market according to the rule

$$p_n^*\left(c_n\right)=\frac{\varepsilon_{q_n}\left(p_n^*\left(c_n\right)\right)}{\varepsilon_{q_n}\left(p_n^*\left(c_n\right)\right)-1}c_n,$$

where $\varepsilon_{q_n}\left(p\right)$ is the price elasticity of the demand function q_n at the price p. Incorporating the optimal pricing rule, we define the variable profits in market n earned at the optimal price

$$v_n^*\left(c_n\right)\equiv q_n\left(p_n^*\left(c_n\right)\right)p_n^*\left(c_n\right)-q_n\left(p_n^*\left(c_n\right)\right)c_n$$

as a function of marginal cost c_n.

Special Cases Our framework from Section 2 posits the constant elasticity demand system, which satisfies Assumption 3 with $q_n\left(p_n\right)=Q_n\left(\frac{p_n}{P_n}\right)^{-\sigma_n}$ where Q_n and P_n are market aggregates. The optimal pricing rule is $p_n^\star\left(c_n\right)=\frac{\sigma_n}{\sigma_n-1}c_n$, which implies constant markups over marginal costs.

Assumption 3 is sufficiently general to allow for variable elasticity of demand and thus variable markups. As an illustrative example, we discuss the Pollak [1971] demand system which also satisfies Assumption 3 and has become popular in the literature studying variable markups [see, e.g., Simonovska, 2015, Klenow and Willis, 2016, Arkolakis et al., 2019, Behrens et al., 2020]. The demand function is characterized by

$$
\text{(12)} \qquad q_n(p_n) = \left(\frac{p_n}{P_n^\star}\right)^{-\sigma_n} + \gamma \qquad , \qquad p_n^\star(c_n) = \frac{\sigma_n}{(\sigma_n - 1) + \left(\frac{p_n^\star(c_n)}{\overline{P}_n}\right)^{\sigma_n}} c_n
$$

where $\gamma < 0$ and $\overline{P}_n$ is the market aggregate choke price. The markup is decreasing in the firm's marginal cost.

G.3. Sufficient Conditions for SCD-C

Given Assumptions 1–3, the firm's variable profits across all markets n net of fixed costs is

$$
\pi(\mathcal{L}, z) = \sum_n v_n^*(c_n(\mathcal{L}, z)) - F_i(\mathcal{L})
$$

and thus its CDCP is to maximize this function with respect to the decision set $\mathcal{L}$.

We now derive a sufficient condition for SCD-C. To begin, we use the gradient theorem to write the marginal value of location ℓ as follows:

$$
D_\ell \pi(\mathcal{L}) = \sum_n \int_0^1 \tilde{\boldsymbol{\xi}}_n(\ell)' \nabla_{\boldsymbol{\Theta}} v_n^* \left(g\left(\boldsymbol{\Theta}_n(\mathcal{L} \setminus \{\ell\}) + t\tilde{\boldsymbol{\xi}}_n(\ell), z\right)\right) \mathrm{d}t - F_\ell
$$

where $\nabla_{\boldsymbol{\Theta}}$ is the gradient operator and $\tilde{\boldsymbol{\xi}}_n(\ell)$ is the $K \times 1$ vector with kth entyr $\tilde{\xi}_{k\ell n}$ which represents the marginal contributions of location ℓ to the production index of each dimension k. Overall, the marginal value of a location ℓ represents the gain in variable profits from increasing each dimension of the index function $\boldsymbol{\Theta}_n$, offset by the additional fixed costs incurred.

The SCD-C condition requires that marginal value only cross zero once. It is sufficient to show the marginal value is monotonic, i.e. given any $\mathcal{L}_1 \subset \mathcal{L}_2 \subseteq L$, the marginal value of any given item ℓ is bigger at $\mathcal{L}_2$ than at $\mathcal{L}_1$ for SCD-C from below, and smaller for SCD-C from above. Comparing this marginal value across the two decision sets,

$$
\begin{aligned}
D_\ell \pi(\mathcal{L}_2, z) = D_\ell \pi(\mathcal{L}_1, z) + \sum_n \int_0^1 \int_0^1 \tilde{\boldsymbol{\xi}}_n(\ell)' H_{\boldsymbol{\Theta}} v_n^* \big(g\big(\boldsymbol{\Theta}_n(\mathcal{L}_1 \setminus \{\ell\}) + t\tilde{\boldsymbol{\xi}}_n(\ell) \\
+ u\boldsymbol{\Delta}(\mathcal{L}_2 \setminus \{\ell\}, \mathcal{L}_1 \setminus \{\ell\})\big), z\big) \boldsymbol{\Delta}_n(\mathcal{L}_2 \setminus \{\ell\}, \mathcal{L}_1 \setminus \{\ell\}) \, \mathrm{d}u \mathrm{d}t
\end{aligned}
$$

where H is the Hessian operator and $\boldsymbol{\Delta}_n(\mathcal{L}_1, \mathcal{L}_2) \equiv \boldsymbol{\Theta}_n(\mathcal{L}_2) - \boldsymbol{\Theta}_n(\mathcal{L}_1)$ is the $K \times 1$ difference between the vector of indices at $\mathcal{L}_1$ and $\mathcal{L}_2$. Since each index is a sum of marginal effects, every entry in the difference $\boldsymbol{\Delta}_n(\mathcal{L}_1, \mathcal{L}_2)$ is positive.

Then, if all elements in the Hessian are positive all n, the difference is guaranteed to be positive and the firm's problem exhibits monotone complements, which is sufficient for SCD-C from below. On the other hand, if all elements of the Hessian are negative for all n, the difference is guaranteed to be negative and the firm's problem exhibits monotone substitutes, which is sufficient for SCD-C from above. Translating this condition to restrictions on the cost and demand functions, the (k, k')th element of the Hessian $H_{\boldsymbol{\Theta}} v_n^*$ is as follows.

$$\frac{\partial^2 v_n^*}{\partial \Theta_{kn} \partial \Theta_{k'n}} = \underbrace{\frac{\partial v_n^*(c)}{\partial c}}_{\equiv v_n^{*\prime}} \underbrace{\frac{\partial g(\boldsymbol{\Theta}_n, z)}{\partial \Theta_{kn}}}_{\equiv g_k'} \left(-\frac{\partial \ln g(\boldsymbol{\Theta}_n, z)}{\partial \Theta_{k'n}} \right) \left\{ \varepsilon_{v_n^{*\prime}} - \frac{\frac{\partial \ln\left[-g_k'(\boldsymbol{\Theta}_n, z)\right]}{\partial \ln \Theta_{k'n}}}{\frac{\partial \ln g(\boldsymbol{\Theta}_n, z)}{\partial \ln \Theta_{k'n}}} \right\}$$

$$\text{where } \varepsilon_{v_n^{*\prime}} \equiv -\frac{\frac{\partial^2 v_n^*(c)}{\partial c^2}}{\frac{\partial v_n^*(c)}{\partial c}} c = -\underbrace{\frac{\mathrm{d} \ln q_n(\mathcal{L}; z)}{\mathrm{d} \ln p_n(\mathcal{L}; z)}}_{\text{Demand Elasticity}} \underbrace{\frac{\mathrm{d} \ln p_n(\mathcal{L}; z)}{\mathrm{d} \ln c_n(\mathcal{L}; z)}}_{\text{Passthrough}}$$

The sign of this element in the Hessian is entirely determined by the term in the curly braces, since the rest is positive by assumption. We summarize the result below.

Proposition (Sufficient condition for SCD-C)**.** *Suppose the firm's problem satisfies Assumptions 1–3. Then, the following condition is sufficient for the firm problem to satisfy SCD-C from below.*

$$(13) \qquad \underbrace{-\frac{\mathrm{d} \ln q_n(\mathcal{L}, z)}{\mathrm{d} \ln p_n(\mathcal{L}, z)} \frac{\mathrm{d} \ln p_n(\mathcal{L}, z)}{\mathrm{d} \ln c_n(\mathcal{L}, z)}}_{\textit{Demand Channel}} \geq \underbrace{\frac{\frac{\partial \ln\left[-g_k'(\boldsymbol{\Theta}_n, z)\right]}{\partial \ln \Theta_{k'n}}}{\frac{\partial \ln g(\boldsymbol{\Theta}_n, z)}{\partial \ln \Theta_{k'n}}}}_{\textit{Supply Channel}} \qquad \forall n, k, k'$$

Reversing the inequality yields a sufficient condition for SCD-C from above.

This condition collapses in the following special cases.

1. *In the case of CES demand, the demand channel collapses to σ_n.*
2. *In the case of Pollak [1971] demand, the demand channel collapses to*

$$\underbrace{\frac{\sigma_n}{1 - \left(\frac{p_n^\star}{P_n}\right)^{\sigma_n}}}_{\textit{Demand Elasticity}} \underbrace{\frac{(\sigma_n - 1) + \left(\frac{p_n^\star}{P_n}\right)^{\sigma_n}}{(\sigma_n - 1) + (\sigma_n + 1)\left(\frac{p_n^\star}{P_n}\right)^{\sigma_n}}}_{\textit{Passthrough}}$$

which is bounded below by σ_n.

3. *In the single-dimensional cost formulation of Tintelnot [2017], Antràs et al. [2017], Arkolakis et al. [2018], the supply channel collapses to ε.*

4. *In the multi-dimensional formulation of Lind and Ramondo [2023], the condition collapses as follows. For all $\mathcal{L}$ and k,*

$$-\frac{\mathrm{d}\ln q_n(\mathcal{L},z)}{\mathrm{d}\ln p_n(\mathcal{L},z)}\frac{\mathrm{d}\ln p_n(\mathcal{L},z)}{\mathrm{d}\ln c_n(\mathcal{L},z)} \leq 1+\theta$$

for SCD-C from above; and

$$-\frac{\mathrm{d}\ln q_n(\mathcal{L},z)}{\mathrm{d}\ln p_n(\mathcal{L},z)}\frac{\mathrm{d}\ln p_n(\mathcal{L},z)}{\mathrm{d}\ln c_n(\mathcal{L},z)} \geq 1+\theta+\frac{\rho_k}{\frac{1-\rho_k}{\theta}\frac{\Theta_{kn}(\mathcal{L})^{1-\rho_k}}{\sum_{k'}\Theta_{k'n}(\mathcal{L})^{1-\rho_{k'}}}}$$

for SCD-C from below.

By assumption, an additional production location always lowers the marginal cost of the firm to supply its final good to any location. Locations are complements in the firm problem when an additional location leads to a larger profit gain the more locations the firm operates; if the reverse, locations are substitutes. Equation (13) decomposes this effect into a supply-side component and a demand-side component.

The supply-side component captures how much an additional production location reduces the marginal cost of the firm, while the demand-side component captures by how much variable profits increase when the marginal cost of the firm drops. The balance of these two forces determines whether location act as complements or substitutes in the firm's overall profit maximization problem. The strength of the demand-side channel depends on the product of the demand and passthrough elasticity. It summarizes the elasticity of variable profits to a change in marginal cost, which is determined by how much a marginal cost change affects the price (passthrough) and in turn by how much demand responds to a marginal decrease in price (demand elasticity).

G.4. Sufficient Conditions for SCD-T

To derive a sufficient condition for SCD-T, we introduce a final Assumption on the role of productivity.

Assumption 4 (Hicks-neutral productivity)**.** The outer function $g:\mathbb{R}^K\times\mathbb{R}^+\rightarrow\mathbb{R}^+$ is multiplicatively separable, so that it can be written

$$g(\mathbf{\Theta},z)=\frac{1}{z}\tilde{g}(\mathbf{\Theta})\ .$$

Following Assumption 4, the difference in profits earned with $\mathcal{L}_1$ compared to $\mathcal{L}_2$ is as follows.

$$\begin{aligned}\pi\left(\mathcal{L}_2,z\right)-\pi\left(\mathcal{L}_1,z\right)=&\sum_n\int_0^1 v_n^{*\prime}\left(\frac{\tilde{g}\left(\boldsymbol{\Theta}_n\left(\mathcal{L}_1\right)+t\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)\right)}{z}\right)\frac{1}{z}\\&\times\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)'\nabla_{\boldsymbol{\Theta}}\tilde{g}\left(\boldsymbol{\Theta}_n\left(\mathcal{L}_1\right)+t\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)\right)\mathrm{d}t\\&-\sum_{\ell\in\mathcal{L}_2}F_\ell+\sum_{\ell\in\mathcal{L}_1}F_\ell\end{aligned}$$

For SCD-T, it is sufficient for this difference to be increasing in z for all pairs that differ by exactly one element, $\{\mathcal{L}_1,\mathcal{L}_2\}=\{\mathcal{L}\setminus\{\ell\},\mathcal{L}\cup\{\ell\}\}$. For strong SCD-T, it is sufficient for this difference to be monotonic in z for all pairs $\{\mathcal{L}_1,\mathcal{L}_2\}$.

We thus derive how this profit difference changes in the productivity z as follows.

$$\begin{aligned}\frac{\partial\left[\pi\left(\mathcal{L}_2,z\right)-\pi\left(\mathcal{L}_1,z\right)\right]}{\partial z}=&\sum_n\int_0^1\left[\varepsilon_{v_n^{*\prime}}-1\right]\frac{v_n^{*\prime}\left(\frac{\tilde{g}\left(\boldsymbol{\Theta}_n\left(\mathcal{L}_1\right)+t\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)\right)}{z}\right)}{z^2}\\&\times\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)'\nabla_{\boldsymbol{\Theta}}\tilde{g}\left(\boldsymbol{\Theta}_n\left(\mathcal{L}_1\right)+t\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)\right)\mathrm{d}t\end{aligned}$$

When $\mathcal{L}_2=\mathcal{L}_1\cup\{\ell\}$, as in the SCD-T condition, all entries of $\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)$ are non-negative with at least one strictly positive, so $\varepsilon_{v_n^{*\prime}}>1$ is sufficient for the marginal value to be monotonically increasing in z. When $\{\mathcal{L}_1,\mathcal{L}_2\}$ is an arbitrary pair of sets, as in the strong SCD-T condition, if all entries in $\boldsymbol{\Delta}_n\left(\mathcal{L}_1,\mathcal{L}_2\right)$ share the same sign and $\varepsilon_{v_n^{*\prime}}>1$, then the derivative never changes sign and the difference is strictly monotonic in z.

Proposition (Sufficient condition for SCD-T)**.** *Suppose the firm's problem satisfies Assumptions 1–4.*

1. *The following condition is sufficient for the problem to satisfy SCD-T.*

$$-\frac{\mathrm{d}\ln q_{in}\left(\mathcal{L},z\right)}{\mathrm{d}\ln p_{in}\left(\mathcal{L},z\right)}\frac{\mathrm{d}\ln p_{in}\left(\mathcal{L},z\right)}{\mathrm{d}\ln c_{in}\left(\mathcal{L},z\right)}>1\qquad\forall n$$

2. *This condition, together with the condition that, for each pair $\{\mathcal{L}_1,\mathcal{L}_2\}$ and markets n, all entries of $\boldsymbol{\Theta}_n\left(\mathcal{L}_2\right)-\boldsymbol{\Theta}_n\left(\mathcal{L}_1\right)$ share the same sign, is sufficient for the problem to satisfy strong SCD-T.*

These conditions collapse in the following special cases.

1. *In the case of CES demand or Pollak [1971] demand, the left hand side collapses as in the sufficient condition for SCD-C.*

2. *In the single-dimensional cost formulation of Tintelnot [2017], Antràs et al. [2017], Arkolakis et al. [2018], the additional condition for strong SCD-T always holds.*

Intuitively, the SCD-T condition requires that the variable profit increase associated with an additional production location is higher at more productive firms, akin to a cross-derivative.

This condition again separates into a demand-side effect on the left and a supply-side effect on the right. The demand-side effect is identical to the one from SCD-C, and describes the elasticity of variable profits to marginal costs. The supply-side effect captures how the reduction in marginal costs associated with an additional production location interacts with firm productivity. An additional production location is worth more at an unproductive firm compared to a productive firm, since the unproductive firm has high marginal costs but can shore up its low productivity by establishing more production locations. In other words, productivity and production sites are substitutes in the firm's cost function. As productivity enters the cost function multiplicatively, the elasticity of substitution between the benefit of a production location and the firm's innate productivity is simply 1.

Under the CES assumption, the condition for SCD-T collapses to $\sigma > 1$. If the production index is also single-dimensional, as in the model of Section 2, then the additional condition for strong SCD-T holds vacuously so $\sigma > 1$ is also sufficient for strong SCD-T.

H. Computational Implementation Details

In this section, we describe the practical implementation of the solution method, as well as the general equilibrium framework which embeds it.

H.1. "Eager" Squeezing

Given the bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, the squeezing step requires computing the marginal value of each location $\ell \in L$ at both the lower and upper bounding decision sets. The computational implementation makes two modifications. First, it only computes the marginal values for locations in $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$. Locations either included in $\underline{\mathcal{L}}$ or excluded from $\overline{\mathcal{L}}$ remain included or excluded, respectively, and need not be rechecked.

Second, the squeezing step is "eager" in the sense that, once an undetermined location is known to be included or excluded, the bounding sets update before computing the marginal value of subsequent undetermined locations. In particular, given an undetermined item $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$, updating occurs as follows.

$$\underline{\mathcal{L}}' = \begin{cases} \underline{\mathcal{L}} & \text{if } D_\ell f\left(\mathcal{L}^{\text{in}}\right) < 0 \\ \underline{\mathcal{L}} \cup \{\ell\} & \text{if } D_\ell f\left(\mathcal{L}^{\text{in}}\right) \geq 0 \end{cases} \qquad \mathcal{L}^{\text{in}} = \begin{cases} \overline{\mathcal{L}} & \text{if SCD-C above} \\ \underline{\mathcal{L}} & \text{if SCD-C below} \end{cases}$$

$$\overline{\mathcal{L}}' = \begin{cases} \overline{\mathcal{L}} \setminus \{\ell\} & \text{if } D_\ell f\left(\mathcal{L}^{\text{out}}\right) < 0 \\ \overline{\mathcal{L}} & \text{if } D_\ell f\left(\mathcal{L}^{\text{out}}\right) \geq 0 \end{cases}, \qquad \mathcal{L}^{\text{out}} = \begin{cases} \underline{\mathcal{L}} & \text{if SCD-C above} \\ \overline{\mathcal{L}} & \text{if SCD-C below} \end{cases}$$

The decision set $\mathcal{L}^{\text{in}}$ is the bounding set that helps determine whether the location is included; similarly, the decision $\mathcal{L}^{\text{out}}$ is the bounding set that helps determine whether the location is excluded.

Eager squeezing implies that, once ℓ is known to be included or excluded, the bounding sets tighten immediately to incorporate this information. Thus, the subsequent undetermined items are considered on the tightened bounding sets To facilitate the eager squeezing, the computational implementation stores an auxiliary set A, which keeps track of the set of locations ℓ which have already been checked with the current bounding sests but have been neither definitely included nor excluded. The squeezing procedure thus has converged once $A = \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$: that is, the marginal value of all undetermined locations has been evaluated at the current bounding sets, and none of them can yet be definitely included or excluded.

H.2. Interval-Based Generalized Squeezing and Refinement

The lower and upper bounding functions imply a partitioning $\mathcal{T}$ on the type space. The computational implementation of the policy function solution explicitly operates on this partitioning. In particular, each interval $\mathcal{Z}_t$ of the partitioning is stored separately as a tuple $\left(\mathcal{Z}_t, \underline{\mathcal{L}}_t, \overline{\mathcal{L}}_t, A_t\right)$, where $\underline{\mathcal{L}}_t$ and $\overline{\mathcal{L}}_t$ are the bounding sets specific to the interval. Then, generalized squeezing refines the partition eagerly, with the auxiliary set A_t tracking the locations in $\overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t$ whose marginal values have been checked at the current bounding sets but have been neither definitively included nor excluded. In particular, given a tuple $\left(\mathcal{Z}_t, \underline{\mathcal{L}}_t, \overline{\mathcal{L}}_t, A_t\right)$, the computational implementation chooses an undetermined location $\ell \in \overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t$ and computes $z_\ell^g\left(\mathcal{L}_t^{\text{in}}\right)$ and $z_\ell^g\left(\mathcal{L}_t^{\text{out}}\right)$. If $z_\ell^g\left(\mathcal{L}_t^{\text{in}}\right)$ is within the interval $\mathcal{Z}_t$, then the partition refines to include ℓ for all $z \in \mathcal{Z}_t$ above this cutoff; similarly, if $z_\ell^g\left(\mathcal{L}_t^{\text{out}}\right)$ is within the interval, then the partition refines to exclude ℓ for all $z \in \mathcal{Z}_t$ below this cutoff. The computational implementation refines each interval independently, and has converged when $A_t = \overline{\mathcal{L}}_t \setminus \underline{\mathcal{L}}_t$ for each interval.

Once generalized squeezing has converged, any interval for which $\underline{\mathcal{L}}_t \neq \overline{\mathcal{L}}_t$ is refined with iterative cutoff search as described in Appendix A, since strong SCD-T holds in our model. In particular, the computational implementation does not use generalized branching.

Finally, the computed policy function is returned as a series of cutoffs $\{z_t\}_{t=0}^{T+1}$ which define the intervals, together with the optimal decision sets for each interval $\{z_t\}_{t=0}^{T}$.

H.3. Aggregation

The general equilibrium conditions require aggregating over the decisions of individual firms. Aggregation in practice is straightforward since the policy function $\mathcal{L}_i^\star(\cdot)$ for firms originating from i is simply a set of productivity intervals $\{[z_{i,t}, z_{i,t+1}]\}_t$ and their associated optimal decision sets $\left\{\mathcal{L}_{i,t}^\star\right\}_t$.

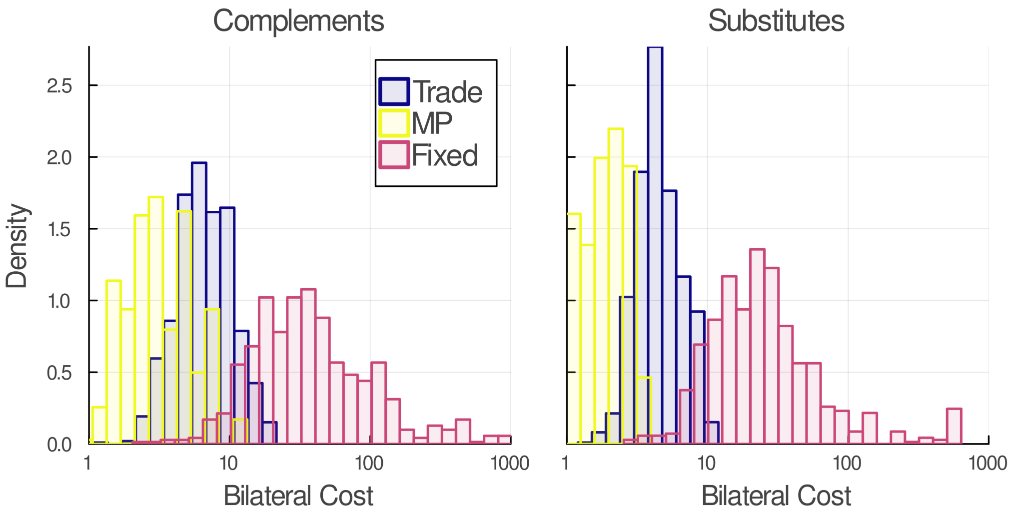
For example, consider the price index equation (5), which requires integrating the pricing decisions across all firms with positive production. Given the computed policy function, the condition can be rewritten as follows.

$$P_n^{1-\sigma} = \sum_i M_i \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} \sum_{\mathcal{Z}_{i,t}\in\mathcal{T}_i} \left(\sum_{\ell\in\mathcal{L}_{i,t}^\star} \xi_{i\ell n}^{1-\varepsilon}\right)^{\frac{\sigma-1}{\varepsilon-1}} \int_{z_{i,t}}^{z_{i,t+1}} z^{\sigma-1} \mathrm{d}G_i(z).$$

In particular, the integral can be divided by the intervals. Since the optimal location set is constant within each interval, integration need only be performed over the firm types. With firm productivity G_i following the Pareto distribution, this integral evaluates closed-form. Aggregation for the other equilibrium conditions follows similarly.

	Complements			Substitutes		
	Trade	MP	Affiliates	Trade	MP	Affiliates
Language	0.061	0.133	0.176	0.044	0.084	0.049
Contiguity	0.165	0.103	0.010	0.145	0.035	0.122
Colony	0.015	-0.060	0.340	0.013	-0.047	0.309
Log Distance	0.283	0.025	0.435	0.225	0.014	0.295

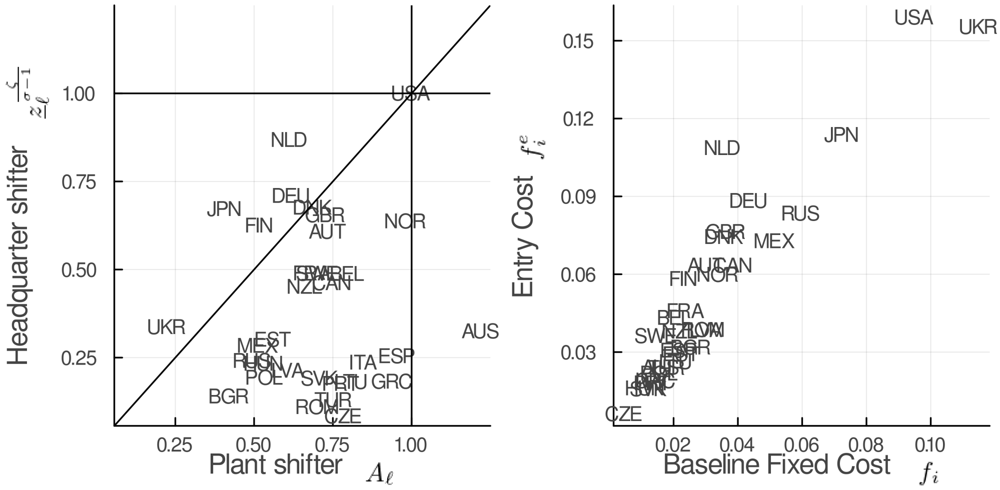
(A) Estimated Cost Elasticities of the Gravity Variables



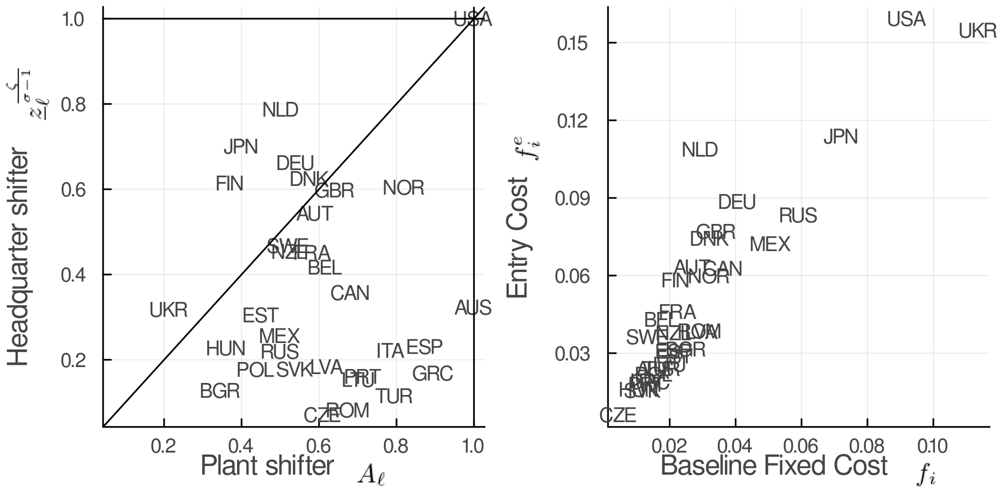
(B) Distribution of Calibrated Bilateral Costs

FIGURE 9: TRADE COSTS, MP COSTS, AND THE BILATERAL COMPONENT OF FIXED COSTS

Note: This figure summarizes the estimated bilateral costs. Table 9a presents the calibrated elasticities of all gravity variables in each of the bilateral costs in the model as specified in equation (11). Figure 9b shows a histogram of the three bilateral cost matrices in the model: trade costs, MP costs, and the bilateral component of fixed costs. We omit the own-country costs which are normalized to 1 for all three types of costs. For MP and the bilateral component of fixed costs, we also omit country pairs where MP is zero, since we set the MP costs to be infinity in those cases.



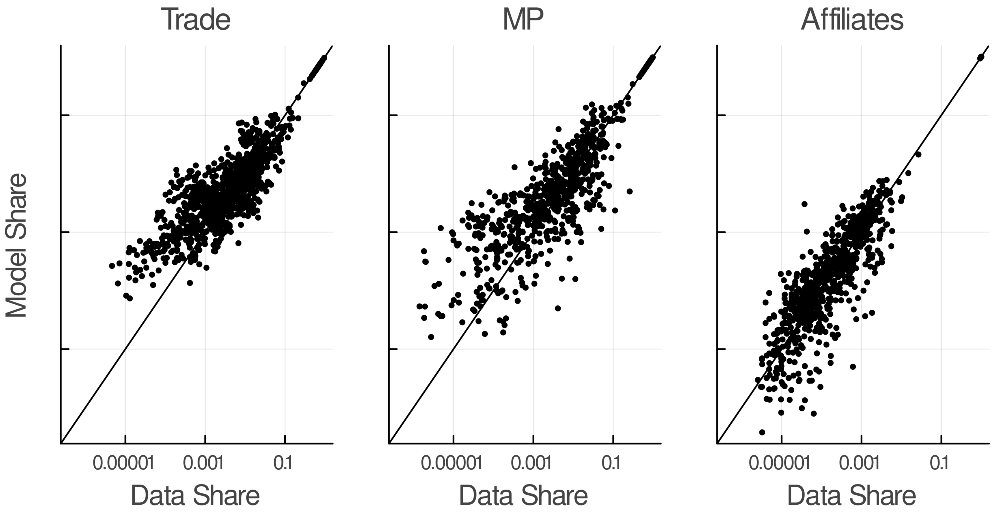
(A) Substitutes



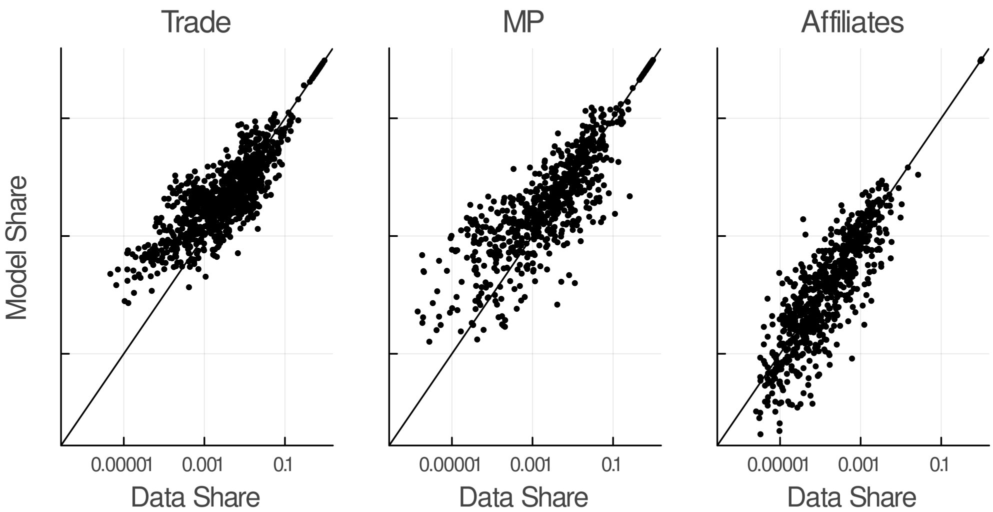
(B) Complements

FIGURE 10: TECHNOLOGY, BASE COMPONENT OF FIXED COSTS, AND ENTRY COSTS IN THE BENCHMARK CALIBRATIONS

Note: The figure shows a number of calibrated shifters in the model. The left panel graphs the Pareto minimum $\underline{z}_i^{\zeta/(\sigma-1)}$ of the firm productivity distribution against the location productivity shifter A_ℓ. The terms $\underline{z}_i^{\zeta/(\sigma-1)}$ and A_ℓ appear multiplicatively in the expression for trilateral flows. The right panel plots the entry cost f_i^e against the base component of the fixed cost f_i.



(A) Substitutes Calibration



(B) Complements Calibration

FIGURE 11: TRADE SHARES, INWARD MP SALES SHARES, AND INWARD AFFILIATE SHARES IN THE DATA AND THE BASELINE CALIBRATIONS

Note: The figure graphs statistics from the data obtained from Alviarez [2019] against the same objects in the calibrated model. The left panel shows trade shares, the second panel shows inward MP sales shares, and the third panel shows inward MP affiliate shares. The correlations between the off-diagonal shares in the model and data are 0.807, 0.774, and 0.839 respectively in the substitutes calibration; and 0.807, 0.765, 0.748 respectively in the complements calibration.

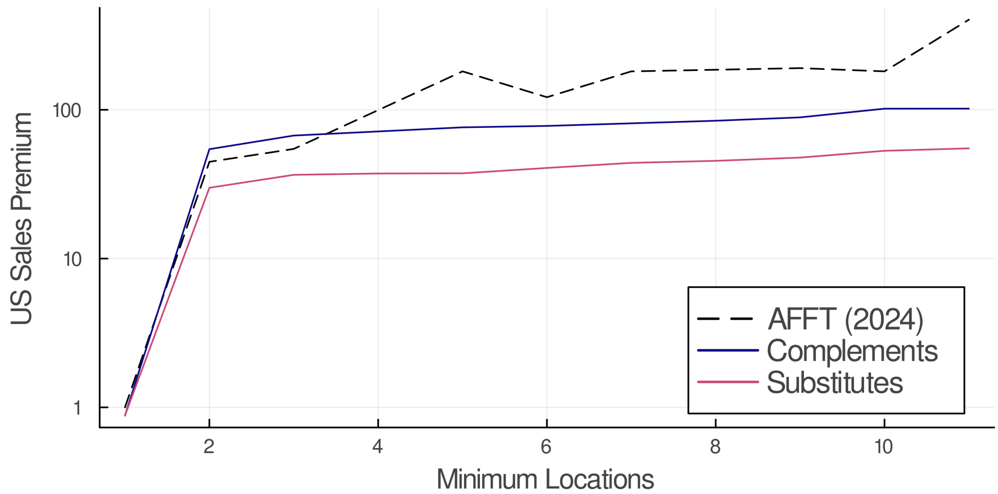
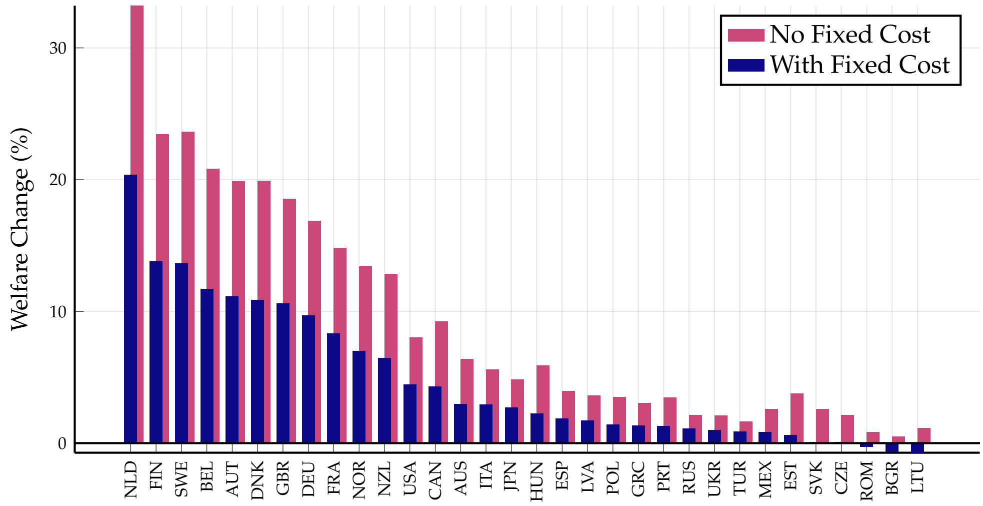
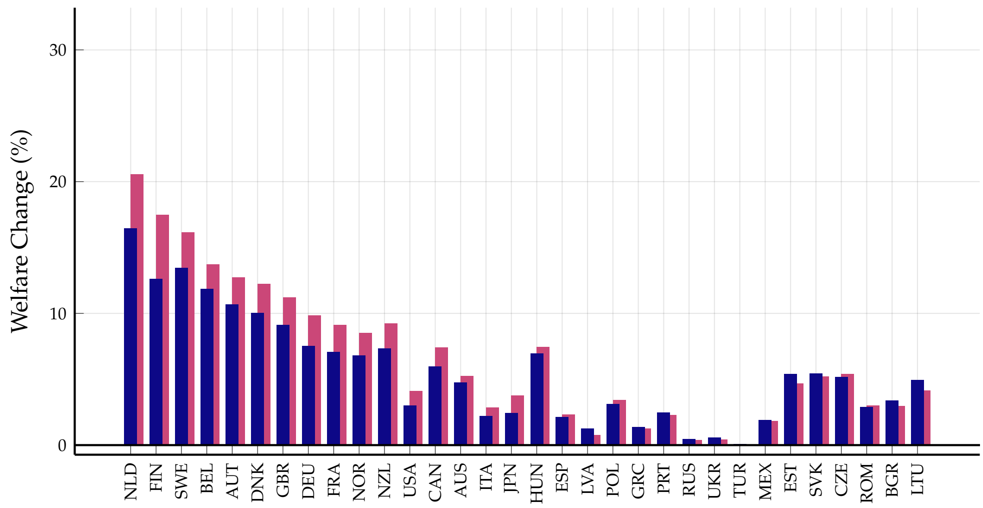


FIGURE 12: MULTINATIONAL SALES PREMIA AND NUMBER OF FOREIGN AFFILIATES IN THE DATA AND THE BASELINE CALIBRATION

Note: The figure compares size sales premia in the model and in the US data obtained from Antràs et al. [2024a]. The sales premium is measured as the relative sales of US-based multinational firms compared to non-MNEs.



(A) Complements



(B) Substitutes

FIGURE 13: WELFARE GAINS OF MULTINATIONAL PRODUCTION WITH AND WITHOUT FIXED COSTS

Note: This figure shows the log point welfare change, $100 \times \ln\left(\hat{w}_i / \hat{P}_i\right)$, from moving from MP autarky to the calibrated economy, for the benchmark level of complementarity and substitutability. The countries are ordered by the size of the welfare effect in the complements calibration with fixed costs.