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 non-empty 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:

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

[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,

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

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

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 $(\mathcal{B}, \supseteq)$. □

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 $y = \sup D \in D$. □

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. □

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 $\mathcal{L}^{\sup} \in E$ be the greatest element. Then, $\overline{\mathcal{L}}^{(K)} \subseteq \mathcal{L}^{\sup}$. Define $\mathcal{L}' = \Phi(\mathcal{L}^{\sup})$ and note that it is also in E. Then, $\mathcal{L}' = \Phi(\mathcal{L}^{\sup}) \subseteq \Phi\left(\overline{\mathcal{L}}^{(K)}\right) = \underline{\mathcal{L}}^{(K)}$ since Φ is order-reversing. Then, $\mathcal{L}' = \underline{\mathcal{L}}^{(K)}$ since $\underline{\mathcal{L}}^{(K)}$ is the smallest element of E, so $\Phi(\mathcal{L}') = \mathcal{L}^{\sup} = \overline{\mathcal{L}}^{(K)}$ 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 $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]$ and an element $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$, define $\tilde{L} \equiv \left(\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}\right) \setminus \{\ell\}$ and the conditional objective function

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

Then, the branching step is defined as

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

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

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

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

Theorem. *Let $\Xi \equiv \{\mathcal{L} \mid \mathcal{L} = \Phi\left(\mathcal{L}\right)\}$ 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 $\left[\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}\right]$ and the one defined by $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}\right]$.

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, $\left(\mathcal{L}' \setminus \underline{\mathcal{L}}\right) \setminus$

$\{\ell\}$ 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

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

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

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

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

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

   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

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

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

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

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 $z < z'$,

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

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

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

or

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

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

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

where $y_{in}(\mathcal{L}, z) = q_n(p_{in}(\mathcal{L}, z))\, p_{in}(\mathcal{L}, z)$ are the firm's total sales in market n and $s_{i\ell n}(\mathcal{L})$ 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 $c_n(\mathcal{L}, z)$ be the unit cost of delivering its final good to a destination market n. Section 2 presents a particular formulation for $c_n(\mathcal{L}, z)$, microfounded by Fréchet location-input-specific cost shocks. Here, we relax the assumption on $c_n(\mathcal{L}, z)$ 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 $c_n(\mathcal{L}) = g(\mathbf{\Theta}_n(\mathcal{L}), z)$ of the vector-valued production index function $\mathbf{\Theta}_n : \mathcal{P}(\mathcal{L}) \to \mathbb{R}^K$ 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 $\mathbf{\Theta}_n(\mathcal{L})$;

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 $\mathbf{\Theta}_n$ 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 $\mathbf{\Theta}_n(\mathcal{L})$ 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

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

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,

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

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

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

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:

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

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,

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

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.

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

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

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.*

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

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 $g : \mathbb{R}^K \times \mathbb{R}^+ \to \mathbb{R}^+$ is multiplicatively separable, so that it can be written

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

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

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

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.

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

When $\mathcal{L}_2 = \mathcal{L}_1 \cup \{\ell\}$, as in the SCD-T condition, all entries of $\boldsymbol{\Delta}_n(\mathcal{L}_1, \mathcal{L}_2)$ 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 $\boldsymbol{\Delta}_n(\mathcal{L}_1, \mathcal{L}_2)$ 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.*

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

2. *This condition, together with the condition that, for each pair $\{\mathcal{L}_1, \mathcal{L}_2\}$ and markets n, all entries of $\boldsymbol{\Theta}_n(\mathcal{L}_2) - \boldsymbol{\Theta}_n(\mathcal{L}_1)$ 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 $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]$, 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 $\left\{\mathcal{L}_{i,t}^\star\right\}_t$.

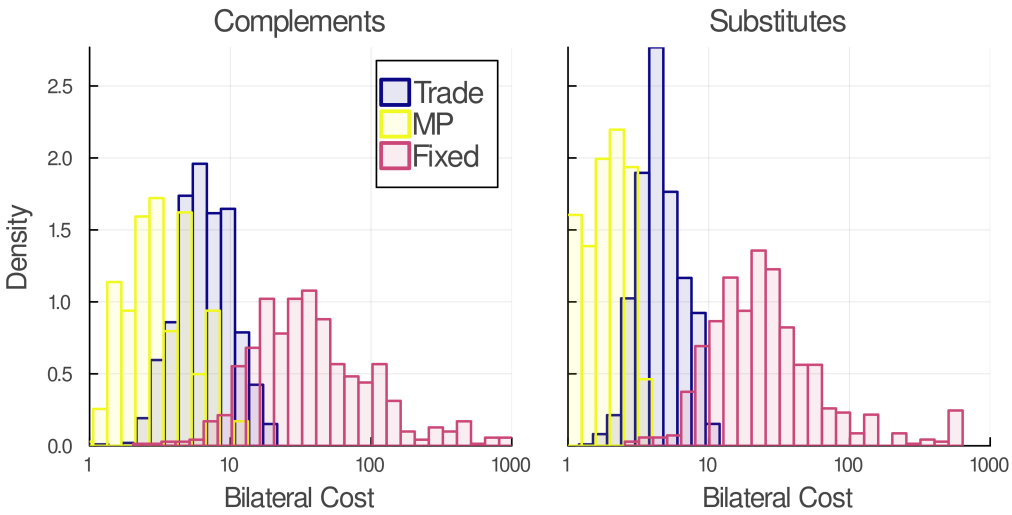
For example, consider the price index equation (5), which requires integrating the pricing decisions across all firms with positive production. Given the computed policy function, the condition can be rewritten as follows.

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

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

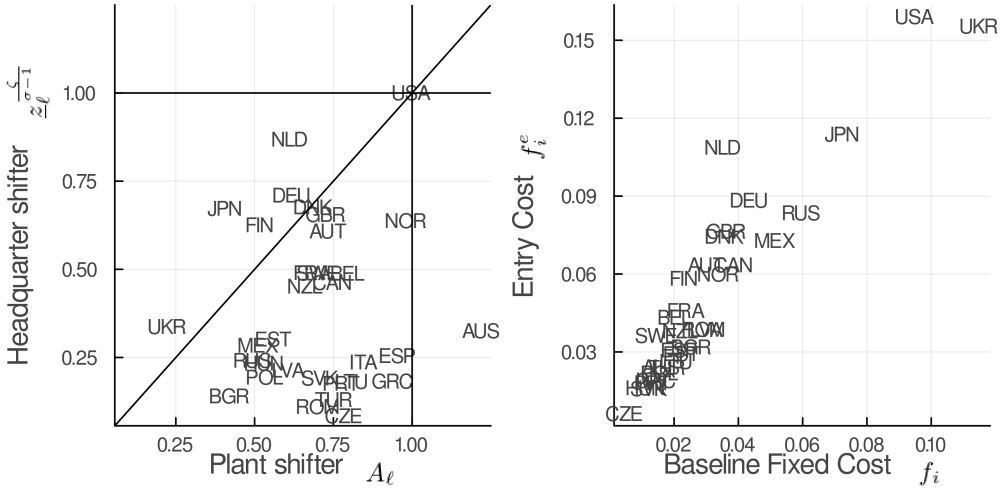
(A) Estimated Cost Elasticities of the Gravity Variables



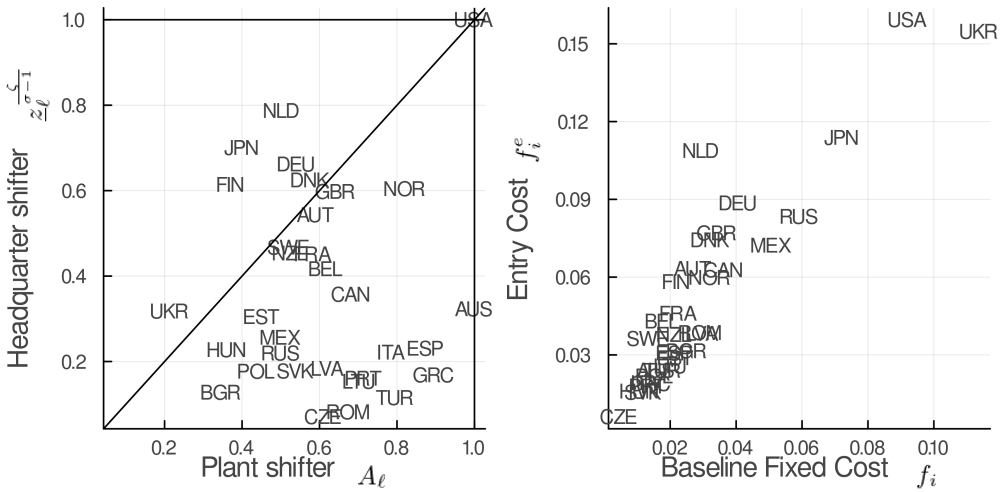
(B) Distribution of Calibrated Bilateral Costs

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.



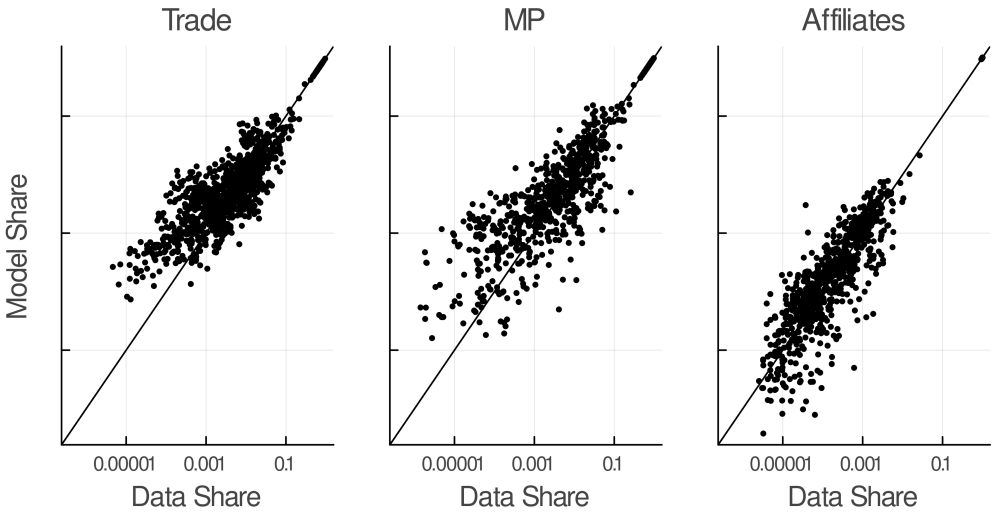
(A) Substitutes



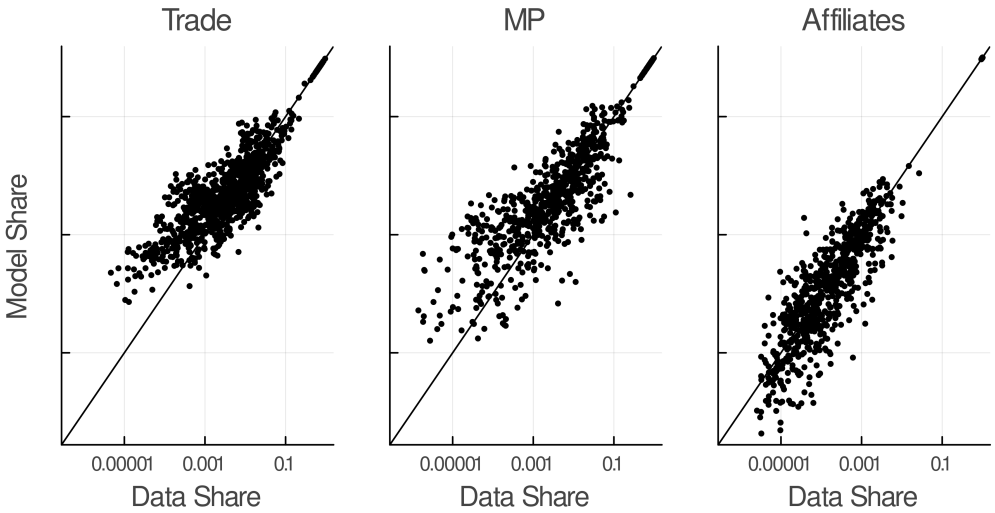
(B) Complements

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.

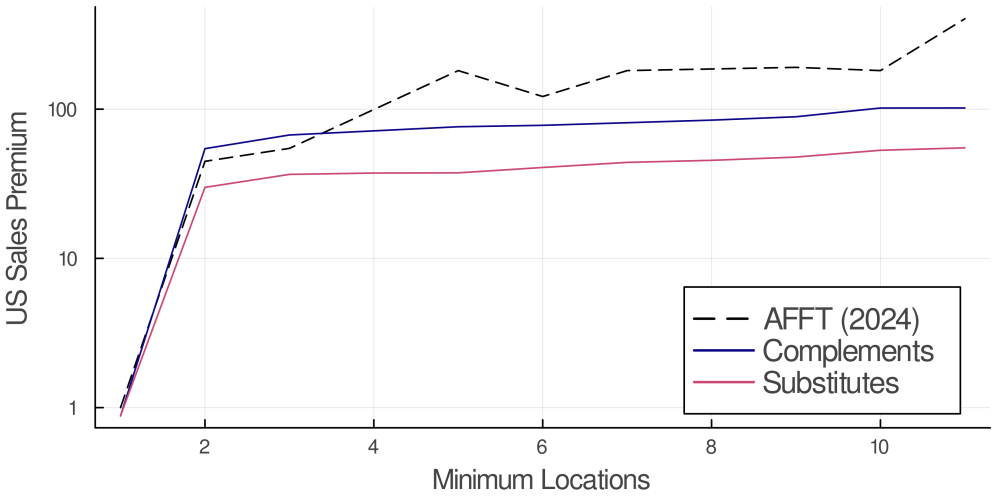
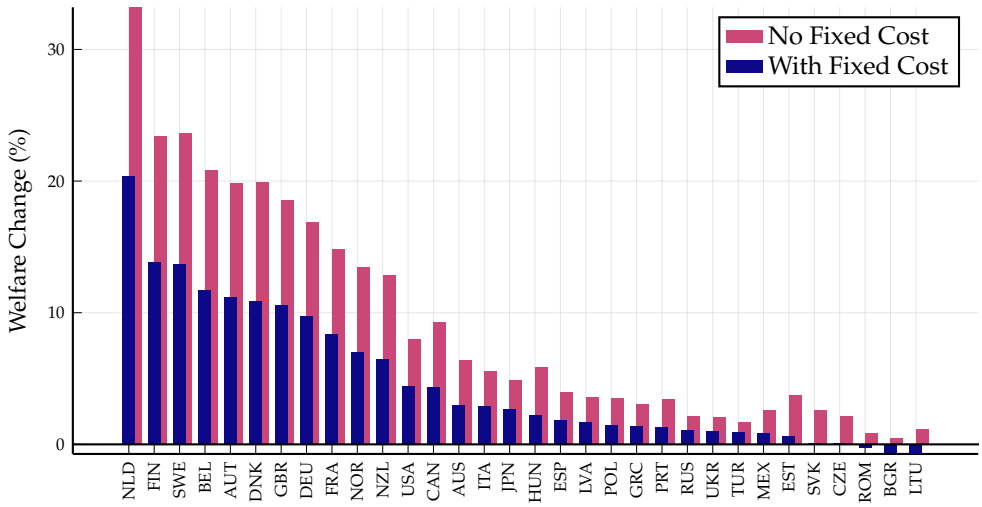
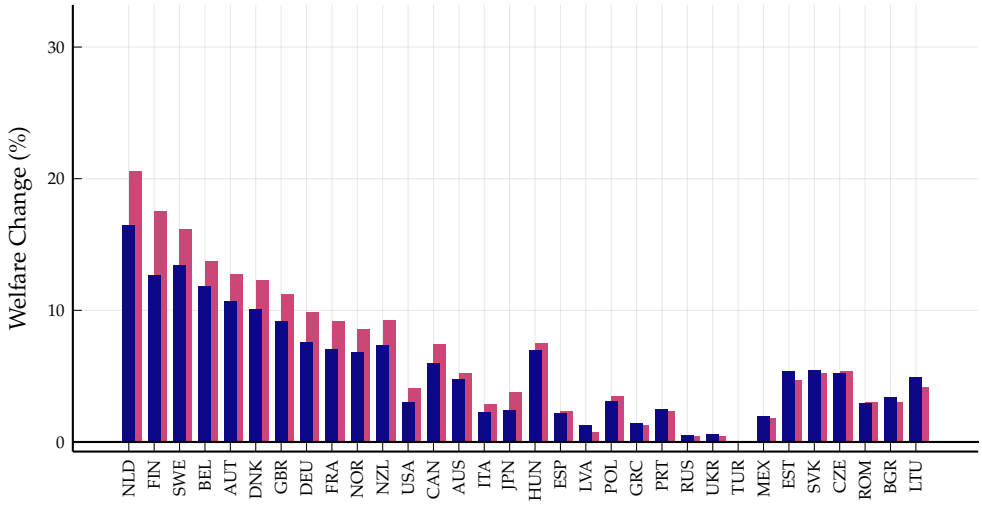


FIGURE 12: MULTINATIONAL SALES PREMIA AND NUMBER OF FOREIGN AFFILIATES IN THE DATA AND THE BASELINE CALIBRATION

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.