ONLINE APPENDIX FOR

COMBINATORIAL DISCRETE CHOICE: A QUANTITATIVE MODEL OF MULTINATIONAL LOCATION DECISIONS

BY COSTAS ARKOLAKIS, FABIAN ECKERT, AND ROWAN SHI

FOR ONLINE PUBLICATION ONLY

OA.1 Additional Theoretical Results

In this section, we provide proofs and derivations for results mentioned in the body of the paper.

OA.1.1 Sufficient Conditions for SCD-C and SCD-T

In this section, we discuss the single crossing difference condition from the main text. We clarify its relationship with the more well-known sub- and super-modularity conditions and the monotone substitutes and complements properties mentioned in the body of the paper. Finally, we verify the sufficient condition for SCD-T provided in the main text.

Proposition 1 (Sufficiency of sub- and super-modularity)**.** *In this first proposition, we show that submodularity and supermodularity are sufficient to ensure SCD-C. Fix the agent type* $\mathbf{z}$ *and consider the return function* π*.*

(1) If π *is submodular, then* π *exhibits SCD-C from above.*

(2) If π *is supermodular, then* π *exhibits SCD-C from below.*

Proof. Since $\mathbf{z}$ is fixed during this proof, we suppress in the following for notational brevity. Begin with a submodular mapping π. Then, by the definition of submodularity, for any sets A, B, it is the case that

$$\pi(A) + \pi(B) \geq \pi(A \cup B) + \pi(A \cap B)\,.$$

We show that SCD-C from above must hold. Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$. Select an arbitrary ℓ. The goal is to show that:

$$\begin{aligned}
\pi(\mathcal{L}_1 \cup \{\ell\}) - \pi(\mathcal{L}_1) &\geq \pi(\mathcal{L}_2 \cup \{\ell\}) - \pi(\mathcal{L}_2) && \text{if } \ell \notin \mathcal{L}_2, \text{ so } \ell \notin \mathcal{L}_1 \\
\pi(\mathcal{L}_1 \cup \{\ell\}) - \pi(\mathcal{L}_1) &\geq \pi(\mathcal{L}_2) - \pi(\mathcal{L}_2 \setminus \{\ell\}) && \text{if } \ell \in \mathcal{L}_2, \text{ but } \ell \notin \mathcal{L}_1 \\
\pi(\mathcal{L}_1) - \pi(\mathcal{L}_1 \setminus \{\ell\}) &\geq \pi(\mathcal{L}_2) - \pi(\mathcal{L}_2 \setminus \{\ell\}) && \text{if } \ell \in \mathcal{L}_1, \text{ so } \ell \in \mathcal{L}_2
\end{aligned}$$

Define the sets A and B as below for each corresponding scenario.

$$\begin{aligned}
&A \equiv \mathcal{L}_1 \cup \{\ell\} && B \equiv \mathcal{L}_2 && \text{if } \ell \notin \mathcal{L}_2, \text{ so } \ell \notin \mathcal{L}_1 \\
&A \equiv \mathcal{L}_1 \cup \{\ell\} && B \equiv \mathcal{L}_2 \setminus \{\ell\} && \text{if } \ell \in \mathcal{L}_2, \text{ but } \ell \notin \mathcal{L}_1 \\
&A \equiv \mathcal{L}_1 && B \equiv \mathcal{L}_2 \setminus \{\ell\} && \text{if } \ell \in \mathcal{L}_1, \text{ so } \ell \in \mathcal{L}_2
\end{aligned}$$

Then, it is easy to see that applying the submodularity condition implies SCD-C from above. Now, suppose π is supermodular. Then, for any sets A, B, it is the case that

$$\pi(A) + \pi(B) \leq \pi(A \cup B) + \pi(A \cap B)\,.$$

We show that SCD-C from below must hold. Let $\mathcal{L}_1 \subseteq \mathcal{L}_2$. Select an arbitrary ℓ. The goal is to show that

$$
\begin{aligned}
\pi(\mathcal{L}_1 \cup \{\ell\}) - \pi(\mathcal{L}_1) &\leq \pi(\mathcal{L}_2 \cup \{l\}) - \pi(\mathcal{L}_2) && \text{if } \ell \notin \mathcal{L}_2, \text{ so } \ell \notin \mathcal{L}_1 \\
\pi(\mathcal{L}_1 \cup \{\ell\}) - \pi(\mathcal{L}_1) &\leq \pi(\mathcal{L}_2) - \pi(\mathcal{L} \setminus \{\ell\}) && \text{if } \ell \in \mathcal{L}_2, \text{ but } \ell \notin \mathcal{L}_1 \\
\pi(\mathcal{L}_1) - \pi(\mathcal{L} \setminus \{\ell\}) &\leq \pi(\mathcal{L}_2) - \pi(\mathcal{L}_2 \setminus \{\ell\}) && \text{if } \ell \in \mathcal{L}_1, \text{ so } \ell \in \mathcal{L}_2
\end{aligned}
$$

Define the sets A and B as above for each corresponding scenario. Then, it is easy to see that applying the supermodularity implies SCD-C from below. □

Next, we show that, in the context of a finite choice space L, the monotone substitutes property implies submodularity, and the monotone complements property implies supermodularity.

Proposition 2. *[Sufficiency for Supermodularity and Submodularity with finite choice space] Fix an agent type $\mathbf{z}$ and consider the return function π. Let A and B be arbitrary sets so that $A \setminus (A \cap B)$ is finite.*
If π exhibits the monotone substitutes property, then

$$\pi(A; \mathbf{z}) + \pi(B; \mathbf{z}) \geq \pi(A \cup B; \mathbf{z}) + \pi(A \cap B; \mathbf{z}) \, .$$

If π exhibits the monotone complements property then

$$\pi(A; \mathbf{z}) + \pi(B; \mathbf{z}) \leq \pi(A \cup B; \mathbf{z}) + \pi(A \cap B; \mathbf{z}) \, .$$

Proof. Since the agent type $\mathbf{z}$ is fixed during this proof, we suppress it for notational brevity in what follows.
Let $\tilde{A}$ and $\tilde{B}$ be arbitrary sets where $\tilde{A} \setminus (\tilde{A} \cap \tilde{B})$ is finite. First, consider the monotone substitutes property. Define

$$I \equiv \tilde{A} \cap \tilde{B} \qquad\qquad A \equiv \tilde{A} \setminus I \qquad\qquad B \equiv \tilde{B} \setminus I \, .$$

Then showing that

$$\pi(\tilde{A}) + \pi(\tilde{B}) \geq \pi(\tilde{A} \cup \tilde{B}) + \pi(\tilde{A} \cap \tilde{B})$$

is equivalent to showing that:

$$\pi(I \cup A) + \pi(I \cup B) \geq \pi(I \cup A \cup B) + \pi(I) \, . \tag{OA.1}$$

The proof proceeds inductively on the cardinality of A. Since $A = \tilde{A} \setminus (\tilde{A} \cap \tilde{B})$, it is finite. When A is empty, then equation (OA.1) holds with equality.

Now suppose equation (OA.1) holds for $|A| = n$. Consider the case where $|A| = n+1$. Let a be an arbitrary element from A and define $\underline{A} \equiv A \setminus \{a\}$. From the inductive assumption,

$$\pi(I) + \pi(I \cup \underline{A} \cup B) \leq \pi(I \cup \underline{A}) + \pi(I \cup B)$$

while from the monotone substitutes property,

$$\begin{aligned} D_a \pi(I \cup \underline{A} \cup B) &\leq D_a \pi(I \cup \underline{A}) \\ \pi(I \cup A \cup B) - \pi(I \cup \underline{A} \cup B) &\leq \pi(I \cup A) - \pi(I \cup \underline{A}) \,. \end{aligned}$$

Combining the two expressions together yields

$$\pi(I) + \pi(I \cup A \cup B) \leq \pi(I \cup A) + \pi(I \cup B) \,,$$

which confirms equation (OA.1) for sets A of cardinality $n+1$. The inductive proof establishes that equation (OA.1) holds for all A of finite size.

Next, consider monotone complements property. The argument follows a similar structure. Now, it is equivalent to show that

$$\pi(I \cup A) + \pi(I \cup B) \leq \pi(I) + \pi(I \cup A \cup B) \,. \tag{OA.2}$$

Proceed inductively once again on the cardinality of A. When A is empty, equation(OA.2) holds with equality. Now suppose equation (OA.2) holds for A with cardinality n. Consider A with cardinality $n+1$. Similarly, select an arbitrary element $a \in A$ and define $\underline{A} \equiv A \setminus \{a\}$. The inductive assumption implies that

$$\pi(I) + \pi(I \cup \underline{A} \cup B) \geq \pi(I \cup \underline{A}) + \pi(I \cup B)$$

while from monotone complements:

$$\begin{aligned} D_a \pi(I \cup \underline{A} \cup B) &\geq D_a \pi(I \cup \underline{A}) \\ \pi(I \cup A \cup B) - \pi(I \cup \underline{A} \cup B) &\geq \pi(I \cup A) - \pi(I \cup \underline{A}) \,. \end{aligned}$$

Combining the two expressions together yields

$$\pi(I) + \pi(I \cup A \cup B) \geq \pi(I \cup A) + \pi(I \cup B) \,,$$

which confirms equation (OA.2) for sets A of cardinality $n+1$. The inductive proof establishes that equation (OA.2) holds for all A of finite size. $\square$

When $A \setminus (A \cap B)$ is not finite, then monotone complements (substitutes) is not sufficient to guarantee supermodularity (submodularity) as the following example shows:

Example. Suppose the return π of a decision set S is defined

$$\pi(S) = \left[\int_S 1 \, \mathrm{d}s\right]^\alpha$$

and note that the marginal value of any item ℓ follows as

$$D_\ell \pi(S) = \left[\int_{S\cup\{\ell\}} 1 \, \mathrm{d}s\right]^\alpha - \left[\int_{S\setminus\{\ell\}} 1 \, \mathrm{d}s\right]^\alpha = 0\,.$$

The intuition is simple: since we integrate over the a decision set S for its return, any singular element ℓ is measure zero and has no effect on the decision set's overall return. The return function therefore satisfies both monotone substitutes and monotone complements. Now consider $A = [0,2]$ and $B = [1,3]$. It is easy to see that

$$\begin{aligned} \pi(A) &= 2^\alpha & \qquad \pi(A \cup B) &= 3^\alpha \\ \pi(B) &= 2^\alpha & \qquad \pi(A \cap B) &= 1^\alpha \end{aligned}$$

so, in this case,

$$\begin{aligned} \alpha > 1 \qquad &\Rightarrow \qquad \pi(A) + \pi(B) > \pi(A \cup B) + \pi(A \cap B) \\ \alpha \in (0,1) \qquad &\Rightarrow \qquad \pi(A) + \pi(B) < \pi(A \cup B) + \pi(A \cap B)\,. \end{aligned}$$

Then, when $\alpha > 1$, the return function obeys monotone substitutes but violates monotone complements. Likewise, when $\alpha \in (0,1)$, the return function obeys monotone complements but violates the monotone substitutes property.

Next, we establish the sufficient condition for SCD-T provided in the main body of the paper. The proof uses again an auxiliary mapping defined in the main text:

$$\Lambda_\ell(\mathcal{L}) = \{\mathbf{z} \in \mathbf{Z} \mid D_\ell(\mathcal{L}; \mathbf{z}) > 0\}\,,$$

which collects all firm efficiency types $\mathbf{z} \in \mathbf{Z}$ for which the marginal value of a given location ℓ is positive given a choice set $\mathcal{L}$.The complement set to $\Lambda_\ell(\mathcal{L})$ is denoted $\Lambda_\ell^c(\mathcal{L})$.

Proposition 3. *[Sufficient condition for SCD-T] Fix an item ℓ and $\mathcal{L}$. Let the entries of $\mathbf{z}$ be indexed by i, so that z_i is the ith coordinate of $\mathbf{z}$. Suppose*

$$\frac{\partial D_\ell \pi(\mathcal{L}; \mathbf{z})}{\partial z_i}$$

(weakly) maintains its sign over the entire type space for each coordinate i. Then, the problem exhibits SCD-T.

Proof. We first show that $\Lambda_\ell(\mathcal{L})$ is a path-connected, and thus connected, set. Let $\mathbf{z}$ and $\mathbf{z}'$ both be in $\Lambda_\ell(\mathcal{L})$. The proof proceeds by constructing a path from $\mathbf{z}$ to $\mathbf{z}'$. First, we construct the point $\tilde{\mathbf{z}}$ where

$$\tilde{z}_i = \begin{cases} \max\{z_i, z_i'\} & \text{if } \frac{\partial D_\ell \pi(\mathcal{L};\mathbf{z})}{\partial z_i} \geq 0 \\ \min\{z_i, z_i'\} & \text{if } \frac{\partial D_\ell \pi(\mathcal{L};\mathbf{z})}{\partial z_i} \leq 0 \end{cases}.$$

Gather the indices $I \equiv \{i \mid z_i \neq \tilde{z}_i\}$. Index them from $m = 1$ to $m = |I|$ and construct the sequence of points $\{\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_m, \ldots, \mathbf{z}_{|I|}\}$ where

$$\mathbf{z}_0 = \mathbf{z} \qquad\qquad \mathbf{z}_m = \mathbf{z}_{m-1} + \mathbf{1}_{i_m}(\tilde{z}_{i_m} - z_{i_m})$$

and $\mathbf{1}_i$ is the ith standard basis vector (that is, the vector with 1 in the ith coordinate and 0 everywhere else). At each step of the sequence, the i_mth coordinate is changed to $\tilde{z}_{i_m}$ and all other coordinates are unchanged.

Then, we construct the piece-wise linear path from $\mathbf{z}$ to $\tilde{\mathbf{z}}$ sequentially passing through these points. This path is contained in $\Lambda_\ell(\mathcal{L})$ by construction. In particular, $D_\ell\pi(\mathcal{L};\cdot)$ starts positive on this path by assumption on $\mathbf{z}$. In each mth segment of the path, only the i_mth component changes while all others stay constant. If the partial derivative of $D_\ell\pi(\mathcal{L};\cdot)$ along this dimension is (weakly) positive, the coordinate is increased; otherwise, it is decreased. Thus, $D_\ell\pi(\mathcal{L};\cdot)$ weakly increases along the path, and so cannot ever fall below zero.

We similarly construct a piece-wise linear path from $\mathbf{z}'$ to $\tilde{\mathbf{z}}$ that lies in $\Lambda_\ell(\mathcal{L})$. Joining these paths together at $\tilde{\mathbf{z}}$, we have constructed a path from $\mathbf{z}$ to $\mathbf{z}'$ that remains in $\Lambda_\ell(\mathcal{L})$. Since $\mathbf{z}$ and $\mathbf{z}'$ were any arbitrary members of $\Lambda_\ell(\mathcal{L})$, we have shown that it is path-connected, and thus connected.

Showing $\Lambda_\ell^c(\mathcal{L})$ is path-connected follows a similar argument. Suppose $\mathbf{z}$ and $\mathbf{z}'$ are contained in $\Lambda_\ell^c(\mathcal{L})$. We construct $\tilde{\mathbf{z}}$ in this case as

$$\tilde{z}_i = \begin{cases} \max\{z_i, z_i'\} & \text{if } \frac{\partial D_\ell \pi(\mathcal{L};\mathbf{z})}{\partial z_i} \leq 0 \\ \min\{z_i, z_i'\} & \text{if } \frac{\partial D_\ell \pi(\mathcal{L};\mathbf{z})}{\partial z_i} \geq 0 \end{cases}.$$

We can then construct the paths from $\mathbf{z}$ to $\tilde{\mathbf{z}}$ and from $\mathbf{z}'$ to $\tilde{\mathbf{z}}$ in the same way. Both lie in $\Lambda_\ell^c(\mathcal{L})$ by the same logic. Therefore, $\Lambda_\ell^c(\mathcal{L})$ is path-connected, and thus connected. $\square$

OA.1.2 Nesting Structure of the Policy Function

The following proposition establishes that, with SCD-C from below, SCD-T, and a single-dimensional type space, the policy function features a nesting structure.

Proposition 4. *[Nested policy function on single dimensional type space] Consider a CDCP (c.f. Definition 1) with single-dimensional type heterogeneity and for which the return function satisfies SCD-C from below and SCD-T. Then, for any $z_1 < z_2$, it must be that $\mathcal{L}^\star(z_1) \subseteq \mathcal{L}^\star(z_2)$.*

Proof. The proof proceeds by contradiction. Suppose $\mathcal{L}^\star(z_1) \nsubseteq \mathcal{L}^\star(z_2)$. Define $\mathcal{L}^o \equiv \mathcal{L}^\star(z_1) \setminus \mathcal{L}^\star(z_2)$, which has cardinality $N \in (0, \infty)$ by assumption. Further, let $\mathcal{L}^i \equiv \mathcal{L}^\star(z_1) \cap \mathcal{L}^\star(z_2)$ so that $\mathcal{L}^\star(z_1) = \mathcal{L}^i \cup \mathcal{L}^o$.

To arrive at a contradiction, we now show that for any finite N, it must be that $\mathcal{L}^\star(z_2) \cup \mathcal{L}^o$ is preferable to $\mathcal{L}^\star(z_2)$ for z_2 types. Note $\mathcal{L}^o \neq \emptyset$, so $\mathcal{L}^\star(z_2) \cup \mathcal{L}^o \neq \mathcal{L}^\star(z_2)$. We proceed by induction on N. Suppose $N = 1$. Index the single element in $\mathcal{L}^o$ by ℓ. Then,

$$
\begin{aligned}
& 0 < D_\ell \pi \left(\mathcal{L}^\star \left(z_1\right), z_1\right) = D_\ell \pi \left(\mathcal{L}^i, z_1\right) \\
\Rightarrow & 0 < D_\ell \pi \left(\mathcal{L}^i, z_2\right) \\
\Rightarrow & 0 < D_\ell \pi \left(\mathcal{L}^\star \left(z_2\right), z_2\right) \\
\Rightarrow & \pi \left(\mathcal{L}^\star \left(z_2\right) \cup \{\ell\}, z_2\right) > \pi \left(\mathcal{L}^\star \left(z_2\right), z_2\right)
\end{aligned}
$$

where the first line derives from z_1 optimality, the second from single-dimensional SCD-T, and the third SCD-C from above, and the fourth from recognizing that $\ell \notin \mathcal{L}^\star(z_2)$. [16] This argument proves the claim for $N = 1$.

Now suppose the claim holds for $N = n$. We now show it holds for $n + 1$. Select any item $\ell \in \mathcal{L}^o$ and let $\mathcal{L}^o_n \equiv \mathcal{L}^o \setminus \{\ell\}$. Then, by the inductive assumption,

$$
\pi(\mathcal{L}^\star(z_2), z_2) < \pi(\mathcal{L}^\star(z_2) \cup \mathcal{L}^o_n, z_2) \,.
$$

Using the optimality of z_1 types, SCD-T, and SCD-C from below,

$$
\begin{aligned}
& 0 < D_\ell \pi \left(\mathcal{L}^\star \left(z_1\right), z_1\right) = D_\ell \pi \left(\mathcal{L}^i \cup \mathcal{L}^o_n; z_1\right) \\
\Rightarrow & 0 < D_\ell \pi \left(\mathcal{L}^i \cup \mathcal{L}^o_n; z_2\right) \\
\Rightarrow & 0 < D_\ell \left(\mathcal{L}^\star(z_2) \cup \mathcal{L}^o_n; z_2\right) \\
\Rightarrow & \pi \left(\mathcal{L}^\star \left(z_2\right) \cup \mathcal{L}^o_n \cup \{\ell\}, z_2\right) > \pi \left(\mathcal{L}^\star \left(z_2\right) \cup \mathcal{L}^o_n, z_2\right) \geq \pi \left(\mathcal{L}^\star \left(z_2\right), z_2\right)
\end{aligned}
$$

where the last inequality makes use of the inductive assumption. Thus we show the claim holds

[16] To break ties, we assume that all items for which an agent is indifferent are excluded from the optimal set.

for $n+1$ if it holds for n.

We have shown that, for any finite non-empty $\mathcal{L}^o$, agents of type z_2 prefer $\mathcal{L}^\star(z_2) \cup \mathcal{L}^o$ to $\mathcal{L}^\star(z_2)$, a contradiction, hence $\mathcal{L}^\star(z_1) \subseteq \mathcal{L}^\star(z_2)$, as claimed. $\square$

The next proposition extends Proposition 4 to the case of multidimensional types $\mathbf{z}$, under a slightly stricter condition on type-location complementarity in place of SCD-T.

Proposition 5. *[Nested policy function on multidimensional type space] Consider a CDCP (c.f. Definition 1) for which the return function satisfies SCD-C from below and for each component i of the type vector* $\mathbf{z}$*,*

$$\frac{\partial D_\ell \pi(\mathcal{L}; \mathbf{z})}{\partial z_i}$$

maintains the same sign for all ℓ*,* $\mathcal{L}$*, and* $\mathbf{z}$*. Consider two types* $\mathbf{z}_1$ *and* $\mathbf{z}_2$ *where the ith component of* $\mathbf{z}_2 - \mathbf{z}_1$ *has the same sign as the ith partial derivative above. Then,* $\mathcal{L}^\star(\mathbf{z}_1) \subseteq \mathcal{L}^\star(\mathbf{z}_2)$*.*

Proof. The proof proceeds by contradiction. Suppose $\mathcal{L}^\star(\mathbf{z}_1) \not\subseteq \mathcal{L}^\star(\mathbf{z}_2)$ so that $\mathcal{L}^o \equiv \mathcal{L}^\star(\mathbf{z}_1) \setminus \mathcal{L}^\star(\mathbf{z}_2) \neq \emptyset$. Let $\mathcal{L}^i \equiv \mathcal{L}^\star(\mathbf{z}_1) \cap \mathcal{L}^\star(\mathbf{z}_2)$ so that $\mathcal{L}^\star(\mathbf{z}_1) = \mathcal{L}^i \cup \mathcal{L}^o$. Observe that, for any convex combination of the two types $\mathbf{z}_1 + \theta(\mathbf{z}_2 - \mathbf{z}_1)$, the marginal value of item j in decision set $\mathcal{L}$ for this type can be expressed using the line integral

$$D_\ell \pi(\mathcal{L}; \mathbf{z}_1 + \theta(\mathbf{z}_2 - \mathbf{z}_1)) = D_\ell \pi(\mathcal{L}; \mathbf{z}_1) + \int_0^\theta \nabla D_\ell \pi(\mathcal{L}; \mathbf{z}_1 + t(\mathbf{z}_2 - \mathbf{z}_1)) \cdot (\mathbf{z}_2 - \mathbf{z}_1) \mathrm{d}t \,.$$

Since the integrand is positive for $\theta > 0$, the integral is positive as well. We may conclude that type $\mathbf{z}_1 + \theta(\mathbf{z}_2 - \mathbf{z}_1)$ receives higher marginal benefit from ℓ's addition in $\mathcal{L}$ than type $\mathbf{z}_1$.

To arrive at a contradiction, we now show that agents of type $\mathbf{z}_2$ prefer $\mathcal{L}^\star(\mathbf{z}_2) \cup \mathcal{L}^o$ to $\mathcal{L}^\star(\mathbf{z}_2)$, a contradiction since $\mathcal{L}^o$ is non-empty. We prove this claim by induction on n the cardinality of $\mathcal{L}^o$. To begin, suppose $|\mathcal{L}^o| = 1$ and let its element ℓ. Then,

$$\begin{aligned}
0 &\leq D_\ell \pi(\mathcal{L}^\star(\mathbf{z}_1), \mathbf{z}_1) \\
&= \pi(\mathcal{L}^i \cup \{\ell\}; \mathbf{z}_1) - \pi(\mathcal{L}^i, \mathbf{z}_1) \\
&\leq \pi(\mathcal{L}^i \cup \{\ell\}, \mathbf{z}_2) - \pi(\mathcal{L}^i; \mathbf{z}_2) \\
&= D_\ell(\mathcal{L}^i; \mathbf{z}_2)
\end{aligned}$$

where the inequality follows from the line integral above. Then, $0 \leq D_\ell(\mathcal{L}^i; \mathbf{z}_2)$, implying that $0 \leq D_\ell(\mathcal{L}^\star(\mathbf{z}_2); \mathbf{z}_2)$ from SCD-C from below. Then, ℓ should optimally be included and $\mathcal{L}^\star(\mathbf{z}_2) \cup \mathcal{L}^o$ is preferred to $\mathcal{L}^\star(\mathbf{z}_2)$. The claim holds for $n = 1$.

Suppose the claim holds for n. To show it must hold for $n+1$, suppose $|\mathcal{L}^o| = n+1$ and select an item from this set to label ℓ. Let $\mathcal{L}^o_n = \mathcal{L}^o \setminus \{\ell\}$. Then,

$$\begin{aligned}
0 &\leq D_\ell \pi(\mathcal{L}^\star(\mathbf{z}_1), \mathbf{z}_1) = D_\ell \pi(\mathcal{L}^i \cup \mathcal{L}^o_n, \mathbf{z}_1) \\
&\leq D_\ell \pi(\mathcal{L}^i \cup \mathcal{L}^o_n, \mathbf{z}_2) \\
0 &\leq D_\ell \pi(\mathcal{L}^\star(\mathbf{z}_2) \cup \mathcal{L}^o_n, \mathbf{z}_2)
\end{aligned}$$

where the second line follows from the line integral and the third from SCD-C from below. We may conclude that an agent with type $\mathbf{z}_2$ prefers $\mathcal{L}^\star(\mathbf{z}_2) \cup \mathcal{L}^o_n \cup \{\ell\}$ to $\mathcal{L}^\star(\mathbf{z}_2) \cup \mathcal{L}^o_n$, which is itself preferred to $\mathcal{L}^\star(\mathbf{z}_2)$ by the inductive assumption.We have shown that, for any finite non-empty $\mathcal{L}^o$, agents of type $\mathbf{z}_2$ prefer $\mathcal{L}^\star(\mathbf{z}_2) \cup \mathcal{L}^o$ to $\mathcal{L}^\star(\mathbf{z}_2)$, a contradiction. Hence $\mathcal{L}^\star(\mathbf{z}_1) \subseteq \mathcal{L}^\star(\mathbf{z}_2)$, as claimed. □

Observe that the condition replacing SCD-T in Proposition 5 is stricter than the sufficiency condition for SCD-T provided in Proposition 3 above. The sufficiency condition allows the ith component of the gradient to differ in its sign for each i, ℓ, and $\mathcal{L}$, as long as it is the same across the type space given these. In contrast, the condition in Proposition 5 requires the sign to remain the same for each i, regardless of which ℓ and $\mathcal{L}$ are chosen.

OA.2 Additional Figures and Tables

Additional Measures of Fit of our Calibrated Model Table OA.1 reports the model fit for moments targeted and untargeted in our calibration. As described in Section 5.2, we are able to exactly match the gravity coefficients and country-level moments we target. Our model also nearly exactly matches both the mean and the standard deviation of trade shares, MP shares, and affiliate shares, the full matrices of which we do not directly target. Table OA.1 also shows that our model exactly replicates the coefficients in Table 1 by adjusting the elasticities of the trade costs, MP costs, and fixed costs to distance, colony, common language, and common border dummies. For all three outcomes, the model explains more than half of the variation in the data. We also report the unconditional correlations for these outcomes between model and data. Overall the fit of our model is very good.

The Determinants of Bilateral Cost Table OA.2 shows the relationship between the calibrated MP costs and fixed costs by distance and other gravity variables. We drop pairs of countries for which we observe no MP activity in the data since we infer infinite MP costs for them. We drop diagonal entries of the cost matrices which are normalized to 1.

The table compares the results from the baseline model on the left against the alternative calibration with no fixed costs. Accordingly, the table on the right does not contain results for fixed costs. Even-numbered columns represent the true specification of costs in the model, reflected in the zero standard errors. On the other hand, odd-numbered columns follow the

Table OA.1: Additional Measures of Model Fit

	Targeted Moments		Untargeted Moments			
			Inward		Outward	
	Data	Model	Data	Model	Data	Model
PPML Gravity Coefficients (Trade, MP, Affiliates)			*Foreign Trade Shares*			
Log Distance	-0.69, -0.29, -0.69	-0.69, -0.29, -0.69	Mean 0.011	0.011	0.010	0.011
Colony	0.08, 0.00, 0.23	0.08, 0.00, 0.23	SD 0.023	0.020	0.023	0.021
Contiguity	0.44, 0.42, 0.45	0.44, 0.42, 0.45	Corr 0.824		0.828	
Language	0.15, 0.47, 0.57	0.15, 0.47, 0.57	R^2 0.678		0.686	
Survival Rate			*Foreign MP Shares*			
Mean	0.86	0.86	Mean 0.009	0.009	0.006	0.006
SD	0.05	0.05	SD 0.026	0.027	0.015	0.015
Corr	1.00		Corr 0.726		0.723	
R^2	1.00		R^2 0.527		0.522	
GDP per Capita (US = 1)			*Foreign Affiliate Shares*			
Mean	0.62	0.62	Mean 0.000	0.000	0.000	0.000
SD	0.21	0.21	SD 0.001	0.001	0.001	0.001
Corr	1.00		Corr 0.757		0.831	
R^2	1.00		R^2 0.573		0.690	

Notes: The left panel shows various moments targeted in our calibration in both the calibrated model and the data. The right panel shows various moments not targeted in the calibration in both calibrated model and data. To compute all moments, we drop the diagonals of all bilateral matrices. While we target the coefficients on gravity variables Table 1, we do not directly target the full matrix of foreign trade, MP, or affiliate shares. The inward shares are such that they sum to 1 if added by destination, across origin. For example, inwards trade shares break down each destination market's imports by import origin. Outwards shares sum to 1 if added by origin, across destination. For example, outwards trade shares break down each production country's exports by export destination.

specification of Alviarez et al. (2023), in particular including origin-specific fixed effects but omitting the colony and contiguity gravity variables.

Additional Counterfactual Figures Figure OA.1 presents the percentage change of affiliate counts in the counterfactual Brexit experiments. Figure OA.2 shows the percentage change of affiliate counts when sanctions are placed on Russia.

OA.3 A More General Theoretical Framework

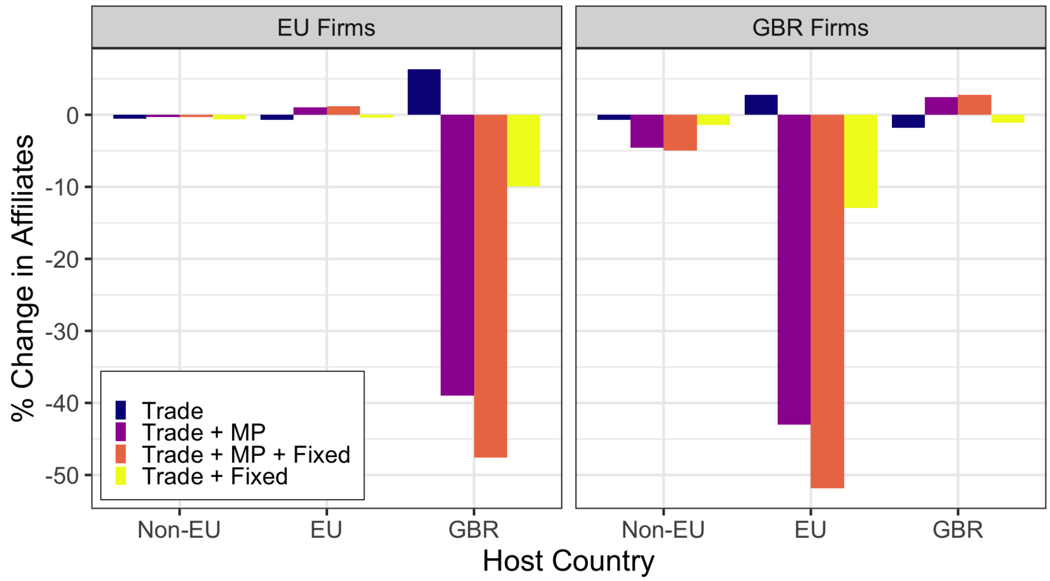
In this section, we relax the assumptions on the production structure and the demand system in our model in Section 2. Instead of specifying a particular production structure, we postulate a cost function that maps a firm's type, its production location set, and its headquarters location

Table OA.2: Calibrated Costs and Gravity Variables

	Baseline Calibration				Without Fixed Costs	
	Log MP Costs		Log Fixed Costs		Log MP Costs	
	(1)	(2)	(3)	(4)	(5)	(6)
Log Distance	0.012^{***}	0.0004^{***}	0.394^{***}	0.354^{***}	0.133^{***}	0.100^{***}
	(0.001)	(0.000)	(0.003)	(0.000)	(0.002)	(0.000)
Colony		0.028^{***}		-0.176^{***}		0.007^{***}
		(0.000)		(0.000)		(0.000)
Contiguity		-0.058^{***}		-0.057^{***}		-0.128^{***}
		(0.000)		(0.000)		(0.000)
Language	-0.061^{***}	-0.065^{***}	-0.155^{***}	-0.082^{***}	-0.140^{***}	-0.127^{***}
	(0.003)	(0.000)	(0.008)	(0.000)	(0.007)	(0.000)
Destination Fe	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	Yes	No	Yes	No	Yes	No
Observations	707	707	707	707	707	707

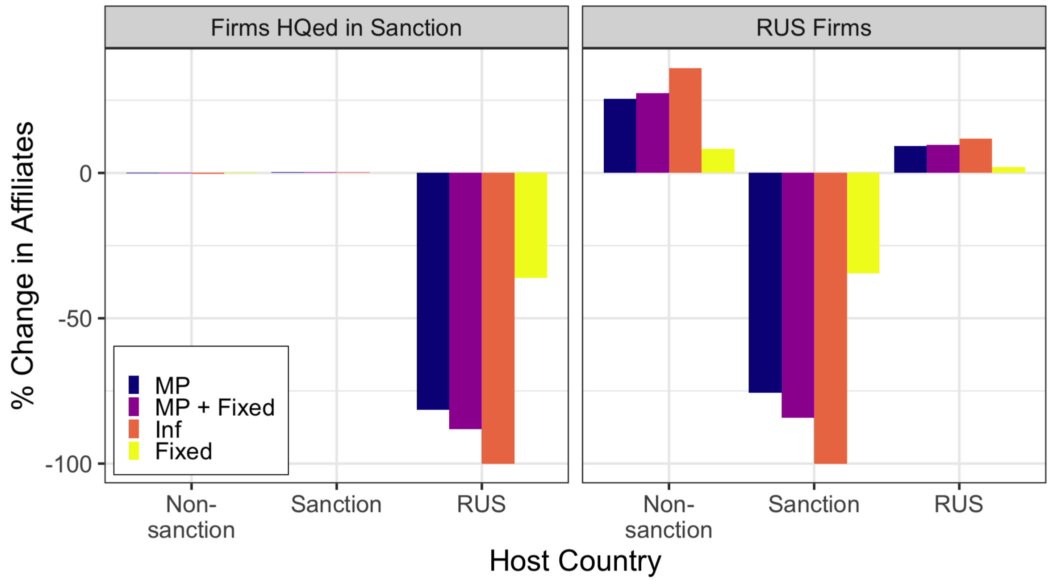
Notes: The table shows results from regressing the log of the calibrated MP costs and fixed costs on log distance. and other gravity variables. The even-numbered columns are the true specification of costs in the model, reflected in the zero standard errors. The odd-numbered columns reflect the specification of Hjort et al. (2022). The left panel shows results from the baseline calibration while the right panel replicates these regressions in the alternative calibration with no fixed costs. We drop country pairs with zero MP for which we infer infinite MP costs and fixed costs. We also drop own country pairs. Levels of significance are denoted as follows: *** Significant at 1 percent level. ** Significant at 5 percent level. * Significant at 10 percent level.

Figure OA.1: Net Percentage Changes in Affiliate Counts by Sender and Host Country in the Brexit Simulation



Notes: The figure shows the percentage change in the number of affiliates operated by firms headquartered in the EU in non-EU countries, EU countries, and in Great Britain in the Brexit simulation. The blue bar reflects a 10% trade costs increase, the purple bar adds a 10% increase of the MP costs, the orange bar adds a 10% increase in fixed costs. The yellow bar considers the trade and fixed cost increases in isolation.

Figure OA.2: Net Percentage Changes in Affiliate Counts by Sender and Host Country in the Sanctions on Russia Simulation



Notes: The first panel of the figure shows the percentage change in the mass of affiliates operated by firms headquartered in countries that placed sanctions to Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries. The ;eft panel of the figure shows the change in the mass of affiliates operated by firms headquartered in Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries. The blue bar refers to a counterfactual with a 30% increase in the cost of MP, the purple bar adds a 30% increase in the fixed costs, the fourth (yellow) bar sets both the MP cost to infinity.

to a unit cost. We then show how to establish single-crossing differences in this more general setup.

OA.3.1 Framework

Cost function Consider a firm of productivity $z \in \mathbb{R}^+$ headquartered in country i with a production location set $\mathcal{L}$ and the unit cost $c_{in}(\mathcal{L}, z)$ of delivering its final good to a destination market n. In Section 2 equation (3), we presented a particular microstructure for $c_{in}(\mathcal{L}, z)$. Instead, in this section, we directly impose structure on $c_{in}(\mathcal{L}, z)$ while remaining agnostic on its microfoundation.

Assumption 1. *The cost function of a firm headquartered in country i with productivity z in destination n can be written as $c_{in}(\mathcal{L}, z) = g(\Theta_{in}(\mathcal{L}), z)$ where $g : \mathbb{R}^2 \to \mathbb{R}^+$ and $\Theta : \mathscr{P} \to \mathbb{R}$, and the production index function Θ is monotonically increasing in the sense that $\mathcal{L} \subseteq \mathcal{L}' \Rightarrow \Theta_{in}(\mathcal{L}) \leq \Theta_{in}(\mathcal{L}')$ for all $\mathcal{L}, \mathcal{L}' \subseteq L$ (if it is decreasing, redefine $\tilde{\Theta}_{in}(\mathcal{L}) = -\Theta_{in}(\mathcal{L})$); and Θ_{in} features no interdependencies among elements of $\mathcal{L}$, i.e., $D_\ell \Theta_{in} \equiv \Theta_{in}(\mathcal{L} \cup \{\ell\}) - \Theta_{in}(\mathcal{L})$ does not depend on $\mathcal{L}$ for all $\ell \in L$ and any $\mathcal{L} \subseteq L$.The "cost function" g is monotonically decreasing in firm productivity (if it's increasing redefine $\tilde{z} \equiv -z$) and in production potential, i.e., $\partial g / \partial z \leq 0$ and $\partial g / \partial \Theta_{in} \leq 0$ hold for all $\mathcal{L} \in L$ and $z \in \mathbb{R}^+$.*

The central object in Assumption 1 is the production "index" function Θ that maps a production location set $\mathcal{L}$ into a scalar. Importantly, this index is such that the marginal contribution of each location to the index is independent of the marginal contribution of other locations.

The cost function $c_{in}(\mathcal{L}, z)$ in equation 3 from our framework in Section 2 satisfies Assumption 1. In particular, in that model, the index function is given as follows:

$$\Theta_{in}(\mathcal{L}) = \sum_{\ell \in \mathcal{L}} \left(\frac{\gamma_{i\ell} w_\ell \tau_{\ell n}}{z_\ell T_\ell} \right)^{-\theta},$$

and the function $g(\cdot)$ by:

$$g(\Theta) = \tilde{\Gamma} \Theta^{-\frac{1}{\theta}}.$$

Many papers in the multinational literature refer to Θ as the "production potential" or "sourcing potential" associated with a given location set (see, e.g., Antras et al. (2017)). 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. In these cases,

$$g_{in}(\Theta, z) = \tilde{\Gamma} \frac{1}{z} \Theta_{in}(\mathcal{L})^{-\frac{1-\rho}{\zeta}}$$

where ρ is the correlation among draws and ζ is the distribution's shape parameter.

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

$$F_i(\mathcal{L}) = \sum_{\ell \in \mathcal{L}} f_{i\ell}.$$

Assumption 2 asserts that there is an independent fixed cost for establishing each production location.

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

Assumption 3. *The firm's variable profit function takes the form*

$$\pi(\boldsymbol{c}) = \sum_{n \in N} \left[q_n(p_n) \, p_n - q_n(p_n) \, c_n \right],$$

where $\mathbf{c} = [c_n]_n$ is the vector of unit costs of producing and delivering a good to the destination markets n.

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 or create strategic considerations among firms. Note we impose little structure on the form of unit costs in this assumption. Instead, we take the vector of unit costs $\mathbf{c}$ as given.

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

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

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

$$\pi_n^*(c_n) \equiv q_n(p_n^*) \, p_n^* - q_n(p_n^*) \, c_n$$

as a function of marginal cost c_n.

OA.3.2 Establishing SCD-C and SCD-T

Throughout this Section, we assume that Assumptions 1-3 hold.[17] We show that there exist easy-to-verify conditions to check whether the firm's problem satisfies SCD-C and SCD-T.

[17] We require that $\partial \pi_{in} / \partial \Theta_{in} \geq 0$ so that profits weakly increase as the value of the index function grows. Note that this always holds under Assumption 1, since $\partial g / \partial \Theta_{in} \leq 0$ implies $\partial \pi_{in} / \partial \Theta_{in} \geq 0$.

Sufficient Conditions for SCD-C We derive a sufficient condition for SCD-C. First notice that given Assumption 1, the marginal contribution of a given production location to the index function Θ_{in} is independent of other production sites. As a result, without loss of generality, we can write Θ_{in} as the sum of the marginal effects of each location on the index:

$$\Theta_{in}(\mathcal{L}) \equiv \sum_{\ell \in \mathcal{L}} \xi_{i\ell n}$$

where the constant $\xi_{i\ell n}$ is the marginal increase in the index from adding ℓ to $\mathcal{L}$ for all $\mathcal{L} \in L$. Next, we can write the marginal value of location j as defined in Definition 2 as follows:

$$D_j \pi_i (\mathcal{L}; z) = \sum_n \left[\pi_n^* \left(c_{in} \left(\mathcal{L} \cup \{j\}; z\right)\right) - \pi_n^* \left(c_{in} \left(\mathcal{L} \setminus \{j\}; z\right)\right)\right] - f_{ij}. \tag{OA.3}$$

The marginal value represents the gain in variable profits from increasing the index function Θ_{in} by ξ_{ijn} which is offset, in part, by the additional fixed costs incurred. The SCD-C condition requires that the expression in equation (OA.3) only crosses zero once. It is sufficient to show the marginal value is monotonic, i.e., given any $\mathcal{L}_1 \subseteq \mathcal{L}_2 \subseteq L$, the marginal value of any given item j is bigger (smaller) at $\mathcal{L}_2$ than for $\mathcal{L}_1$ for SCD-C from below (above). Comparing this marginal value across two decision sets and making use of the fundamental theorem of calculus, we have

$$\begin{aligned}
D_j\pi_i(\mathcal{L}_2; z) - D_j\pi_i(\mathcal{L}_1; z) &= \sum_n \int_0^{\xi_{ijn}} \left[\frac{\partial \pi_n^*}{\partial \Theta_{in}}\left(y + \sum_{\ell \in \mathcal{L}_2} \xi_{i\ell n}, z\right) - \frac{\partial \pi_n^*}{\partial \Theta_{in}}\left(y + \sum_{\ell \in \mathcal{L}_1} \xi_{i\ell n}, z\right)\right] \mathrm{d}y \\
&= \sum_n \int_{\Theta(\mathcal{L}_1)}^{\Theta(\mathcal{L}_2)} \frac{\partial}{\partial x}\left[\int_0^{\xi_{ijn}} \frac{\partial \pi_n^*}{\partial \Theta_{in}}(x+y, z)\mathrm{d}y\right] \mathrm{d}x \\
&= \sum_n \int_{\Theta(\mathcal{L}_1)}^{\Theta(\mathcal{L}_2)} \left[\int_0^{\xi_{ijn}} \frac{\partial^2 \pi_n^*}{\partial \Theta_{in}^2}(x+y, z)\mathrm{d}y\right] \mathrm{d}x
\end{aligned}$$

where we assume the second derivative of the profit function exists and use the fact that, as a direct consequence of the index function being a sum of marginal effects, the value of the index evaluated at $\mathcal{L}_2$ exceeds its value at $\mathcal{L}_1$:

$$\Theta_{in}(\mathcal{L}_1) = \sum_{\ell \in \mathcal{L}_1} \xi_{i\ell n} \leq \sum_{\ell \in \mathcal{L}_2} \xi_{i\ell n} = \Theta_{in}(\mathcal{L}_2).$$

Thus, $\frac{\partial^2 \pi_n^*}{\partial \Theta_{in}^2} \geq 0$ for all markets n is sufficient to guarantee the sign of the RHS is positive, i.e., the marginal value of k in the larger set $\mathcal{L}_2$ exceeds its marginal value in the smaller set $\mathcal{L}_1$. Then, we are guaranteed SCD-C from below. Similarly, $\frac{\partial^2 \pi_n^*}{\partial \Theta_{in}^2} \leq 0$ for all markets n is sufficient to guarantee SCD-C from above.

We next translate this condition to conditions on the demand and the cost function. The second derivative of the destination-specific profit function with respect to the production is given by:

$$\frac{\partial^2 \pi_n^*}{\partial \Theta_{in}^2} = \underbrace{\frac{\partial \pi_n^*(c)}{\partial c}}_{\equiv \pi_n^{*\prime}} \underbrace{\frac{\partial g(\Theta_{in}, z)}{\partial \Theta_{in}}}_{\equiv g_1'} = \frac{\varepsilon_g}{\Theta_{in}} \left[\varepsilon_{\pi_n^{*\prime}} - \frac{\varepsilon_{g_1'}}{\varepsilon_g} \right] \tag{OA.4}$$

where the elasticity of the derivative of the profit function is

$$\varepsilon_{\pi_n^{*\prime}} = -\frac{\frac{\partial^2 \pi_n^*(c)}{\partial c^2}}{\frac{\partial \pi_n^*(c)}{\partial c}} c^2 = \underbrace{\varepsilon_q(p)}_{\text{demand elasticity}} \underbrace{\frac{\mathrm{d} \ln p}{\mathrm{d} \ln c}}_{\text{passthrough}} .$$

The sign of $\frac{\partial^2 \pi_n^*}{\partial \Theta_{in}^2}$ is determined by the component in square brackets since the term pre-multiplying them is positive. All together, the firm's profit function satisfies SCD-C from below if the term in square brackets is positive, i.e., if $\varepsilon_{\pi_n^{*\prime}}$ exceeds $\frac{\varepsilon_{g_1'}}{\varepsilon_g}$, and SCD-C from above if the inequality is reversed.

Similarly to the main text, the condition in equation (OA.4) separates demand and supply side forces. The modeling assumptions in Section 2 implied that demand side forces were always inducing positive complementarities among production locations and supply side forces negative complementarities. For example, it is possible for the cost side term $\frac{\varepsilon_{g_1'}}{\varepsilon_g} < 0$, if the locations are complements in cost. One microfoundation for these cost-side complementarities could be scale economies in the number of production sites, agglomeration in the density of production locations, or complementarities that may arise from location-level specialization.[18]

Sufficient Conditions for SCD-T A similar argument holds for SCD-T in our context, when the productivity is Hicks-neutral, according to

$$c_{in}(\mathcal{L}; z) = \frac{\tilde{\Gamma}}{z} \left[\sum_{\ell \in \mathcal{L}} \xi_{i\ell n} \right]^{-\frac{1}{\theta}} \equiv \frac{\tilde{g}(\mathcal{L})}{z}.$$

The marginal value of a location ℓ to a set $\mathcal{L}$ is

$$\begin{aligned} \sum_n \left[\pi_n^* \left(\frac{\tilde{g}(\mathcal{L} \cup \ell)}{z} \right) - \pi_n^* \left(\frac{\tilde{g}(\mathcal{L})}{z} \right) \right] - f_{i\ell} &= \sum_n \int_{\frac{\tilde{g}(\mathcal{L})}{z}}^{\frac{\tilde{g}(\mathcal{L} \cup \ell)}{z}} \pi_n^{*\prime}(c) \mathrm{d}c - f_{i\ell} \\ &= \sum_n \int_{\tilde{g}(\mathcal{L})}^{\tilde{g}(\mathcal{L} \cup \ell)} \pi_n^{*\prime} \left(\frac{x}{z} \right) \frac{1}{z} \mathrm{d}x - f_{i\ell} \end{aligned}$$

[18] Though it is possible for the demand-side complementarities to be negative, through either negative demand elasticity or passthrough, this scenario is less likely.

via a change in variables. Since $\tilde{g}(\mathcal{L}) > \tilde{g}(\mathcal{L} \cup \ell)$, we have

$$\pi_n^*\left(\frac{\tilde{g}\left(\mathcal{L}\cup\ell\right)}{z}\right) - \pi_n^*\left(\frac{\tilde{g}(\mathcal{L})}{z}\right) - f_{i\ell} = \int_{\tilde{g}(\mathcal{L}\cup\ell)}^{\tilde{g}(\mathcal{L})}\left[-\pi_n^{*\prime}\left(\frac{x}{z}\right)\right]\frac{1}{z}\mathrm{d}x - f_{i\ell}$$

where $\pi_n^{*\prime} \leq 0$ since variable profits decrease in marginal cost. SCD-T requires that this marginal value crosses zero at one productivity value at most. It is sufficient to show that the marginal value is monotonic in productivity. Comparing the marginal value at two values of productivity $z_1 \leq z_2$,

$$\begin{aligned}
&\int_{\tilde{g}(\mathcal{L}\cup\ell)}^{\tilde{g}(\mathcal{L})}\left[-\pi_n^{*\prime}\left(\frac{x}{z_2}\right)\right]\frac{1}{z_2}\mathrm{d}x - \int_{\tilde{g}(\mathcal{L}\cup\ell)}^{\tilde{g}(\mathcal{L})}\left[-\pi_n^{*\prime}\left(\frac{x}{z_1}\right)\right]\frac{1}{z_1}\mathrm{d}x \\
&= \int_{z_1}^{z_2}\frac{\partial}{\partial z}\left\{\int_{\tilde{g}(\mathcal{L}\cup\ell)}^{\tilde{g}(\mathcal{L})}\left[-\pi_n^{*\prime}\left(\frac{x}{z}\right)\right]\frac{1}{z}\mathrm{d}x\right\}\mathrm{d}z \\
&= \int_{z_1}^{z_2}\int_{\tilde{g}(\mathcal{L}\cup\ell)}^{\tilde{g}(\mathcal{L})}\left[-\pi_n^{*\prime}\left(\frac{x}{z}\right) - \frac{\nu}{z}\pi_n^{*\prime\prime}\left(\frac{x}{z}\right)\right]\left(-\frac{1}{z^2}\right)\mathrm{d}x\mathrm{d}z \\
&= \int_{z_1}^{z_2}\int_{\tilde{g}(\mathcal{L}\cup\ell)}^{\tilde{g}(\mathcal{L})}\left[-\frac{1}{z^2}\pi_n^{*\prime}\left(\frac{x}{z}\right)\right]\left[-\frac{\pi_n^{*\prime\prime}\left(\frac{x}{z}\right)}{\pi_n^{*\prime}\left(\frac{x}{z}\right)}\frac{x}{z} - 1\right]\mathrm{d}x\mathrm{d}z \\
&= \int_{z_1}^{z_2}\int_{\tilde{g}(\mathcal{L}\cup\ell)}^{\tilde{g}(\mathcal{L})}\left[-\frac{1}{z^2}\pi_n^{*\prime}\left(\frac{x}{z}\right)\right]\left[\varepsilon_{\pi_n^{*\prime}} - 1\right]\mathrm{d}x\mathrm{d}z
\end{aligned}$$

so it is sufficient to show this difference is positive. Note that the first square bracketed term is positive. Then, it is sufficient for $\varepsilon_{\pi_n^{*\prime}} \geq 1$ for SCD-T to hold. Recall

$$\varepsilon_{\pi_n^{*\prime}} = \varepsilon_q(p)\frac{\mathrm{d}\ln p}{\mathrm{d}\ln c},$$

so SCD-T holds as long as

$$\varepsilon_q(p)\frac{\mathrm{d}\ln p}{\mathrm{d}\ln c} \geq 1.$$

A Multivariate Generalization The previous conditions can be extended to a case where the firm's cost function is instead a multivariate function $g : \mathbb{R}^K \to \mathbb{R}$ that maps K different production indexes Θ to the cost function. In particular, Lind and Ramondo (2023) present a microfoundation for such a cost function, where

$$g(\mathbf{\Theta}; z) = \frac{\tilde{\Gamma}}{z}\left[\sum_k \Theta_{kin}(\mathcal{L})^{1-\rho_k}\right]^{-\frac{1}{\theta}} \quad , \quad \Theta_{kin}(\mathcal{L}) = \sum_{\mathcal{L}} \xi_{ki\ell n}$$

derives from a nested multivariate Fréchet structure with correlation ρ_k within nests indexed by k, and the contribution $\xi_{ki\ell n}$ of location ℓ to serving destination n as a firm based in i can vary by nest k. The terms $\xi_{ki\ell n}$ are constant across firms, with similar interpretation within nests to $\xi_{i\ell n}$ in the baseline model. Finally, $\tilde{\Gamma}$ is a constant of integration. Correlated nests, as discussed in Lind and Ramondo (2023), can arise from models with K different technological techniques of production, industries, or other latent variables.

With this structure, the marginal value of item j now affects each of the K indices $\Theta_{kin}(\mathcal{L} \cup \{j\})$ for each nest. In particular, using the gradient theorem,

$$\begin{aligned}
D_j \pi_i (\mathcal{L}_1; z) &= \sum_n \left[\pi_n^* \left(c_{in} \left(\mathcal{L}_1 \cup \{j\}; z\right)\right) - \pi_n^* \left(c_{in} \left(\mathcal{L}_1 \setminus \{j\}; z\right)\right)\right] - f_{ij} \\
&= \sum_n \int_0^1 \sum \boldsymbol{\xi}_{in}(j)' \left[\nabla_{\boldsymbol{\Theta}} \pi_n^* \left(\boldsymbol{\xi}_{in}(j) t + \boldsymbol{\Theta}\left(\mathcal{L}_1\right); z\right)\right] \mathrm{d}t - f_{ij}
\end{aligned}$$

where we let $\boldsymbol{\xi}_{in}(j)$ be the $K \times 1$ vector with kth element ξ_{kijn}.

To derive a sufficient condition for SCD-C, we compare the marginal value of j to the sets $\mathcal{L}_1$ and $\mathcal{L}_2$ for any pair $\mathcal{L}_1 \subseteq \mathcal{L}_2$,

$$\begin{aligned}
D_j \pi_i (\mathcal{L}_2) - D_j \pi_i (\mathcal{L}_1) = \sum_n \int_0^1 & \boldsymbol{\xi}_{in}(j)' \nabla_{\boldsymbol{\Theta}} \left[\pi_n^* \left(\boldsymbol{\xi}_{in}(j) t + \boldsymbol{\Theta}\left(\mathcal{L}_2\right); z\right)\right. \\
& \left. - \pi_n^* \left(\boldsymbol{\xi}_{in}(j) t + \boldsymbol{\Theta}\left(\mathcal{L}_1\right); z\right)\right] \mathrm{d}t \\
= \sum_n \int_0^1 & \boldsymbol{\xi}_{in}(j)' \nabla_{\boldsymbol{\Theta}} \int_0^1 \nabla_{\boldsymbol{\Theta}} \pi_n^* \left(\boldsymbol{\Theta}\left(\mathcal{L}_1\right)\right. \\
& \left. + \boldsymbol{\xi}_{in}(j) t + \left(\boldsymbol{\Theta}\left(\mathcal{L}_2\right) - \boldsymbol{\Theta}\left(\mathcal{L}_1\right)\right) \oslash \boldsymbol{\xi}_{in}(j) \odot \boldsymbol{\xi}_{in}(j) u; z\right) \boldsymbol{\xi}_{in}(j) \mathrm{d}u \mathrm{d}t \\
= \sum_n \int_0^1 & \int_0^1 \boldsymbol{\xi}_{in}(j)' H_{\boldsymbol{\Theta}} \pi_n^* \left(\boldsymbol{\Theta}\left(\mathcal{L}_1\right)\right. \\
& \left. + \boldsymbol{\xi}_{in}(j) t + \left(\boldsymbol{\Theta}\left(\mathcal{L}_2\right) - \boldsymbol{\Theta}\left(\mathcal{L}_1\right)\right) \oslash \boldsymbol{\xi}_{in}(j) \odot \boldsymbol{\xi}_{in}(j) u; z\right) \boldsymbol{\xi}_{in}(j) \mathrm{d}u \mathrm{d}t
\end{aligned}$$

where $\odot$ and $\oslash$ denote Hadamard (element-wise) multiplication and division, respectively, and H is the Hessian operator. We derive the second line in the above by applying the gradient theorem a second time.

Then, if the Hessian of π_n^* is positive semidefinite for all n, the difference is guaranteed to be positive and the firm's problem exhibits monotone complements (sufficient for SCD-C from below). On the other hand, if the Hessians for all n is negative semidefinite, the difference is guaranteed to be negative and the firm's problem exhibits monotone substitutes (sufficient for SCD-C from above).

The Hessian of the function π_n^* derived with respect to $\boldsymbol{\Theta}$ has as its (k, k')th element is a generalization of the single-dimensional case, as follows,

$$\frac{\partial \pi_n^*(c)}{\partial c} \frac{\partial g\left(\boldsymbol{\Theta}, z\right)}{\partial \Theta_{kin}} \left(-\frac{\partial \ln g\left(\boldsymbol{\Theta}, z\right)}{\partial \Theta_{k'in}}\right) \left[\varepsilon_{\pi_n^{*\prime}} - \frac{-\frac{\partial \ln g_k'(\boldsymbol{\Theta}, z)}{\partial \ln \Theta_{k'in}}}{-\frac{\partial \ln g(\boldsymbol{\Theta}, z)}{\partial \ln \Theta_{k'in}}}\right]$$

where g_i' is the partial derivative of g with respect to the kth element of $\boldsymbol{\Theta}$. Similarly to the single dimensional case, the terms preceding the square brackets are positive. Thus, the sign of each element in the Hessian is determined entirely by the term in the square brackets. Thus, an extremely strong sufficient condition for monotone complements is for the term in the square brackets to be positive for all n and nest pairs (k, k'). Similarly, an extremely strong sufficient condition for monotone substitutes is for the term in the square brackets to be negative for all n and nest pairs (k, k').

As an example, consider the multivariate g function in Lind and Ramondo (2023). To derive the sign of the term in the square brackets, note that $\varepsilon_{\pi_n^{*\prime}}$ remains the product of passthrough and the elasticity of demand, since we do not change our assumption on the structure of π_n^*. However, computing the terms deriving from the cost side,

$$\frac{-\frac{\partial \ln g_k'(\mathbf{\Theta}, z)}{\partial \ln \Theta_{k'in}}}{-\frac{\partial \ln g(\mathbf{\Theta}, z)}{\partial \ln \Theta_{k'in}}} = \begin{cases} 1+\theta & \text{if } k \neq k' \\ 1+\theta+\frac{\rho_k}{\frac{1-\rho_k}{\theta}\frac{\Theta_{kin}^{1-\rho_k}}{\sum_{k'}\Theta_{k'in}^{1-\rho_{k'}}}} & \text{if } k = k' \end{cases}.$$

Thus, $\sigma \leq 1+\theta$ is more than sufficient to guarantee that the firm's problem will display the monotone substitutes property, and thus SCD-C from above. Intuitively, this cost function introduces correlation across the Fréchet draws within the same nest k, but no correlation across nests. Thus, when we consider two different nests $k \neq k'$, we derive the original condition for monotone substitutes that we had from the uncorrelated Fréchet distribution. However, if we consider the behavior within the same nest, $k = k'$, the draws are correlated. In this case, they become more closely substitutable, and thus the "cannibalization" effect on the cost side grows larger than the $1+\theta$ benchmark from the uncorrelated case.

The nested correlation structure precludes easily testing for monotone complements. However, as long as $\frac{\Theta_{kin}^{1-\rho_k}}{\sum_{k'}\Theta_{k'in}^{1-\rho_{k'}}}$, the relative importance of nest k among all nests, can be bounded from below, then the cost-side cannibalization effect can be bounded from above. In this case, it is sufficient for the demand-side complementarities to exceed the upper bound of the cost-side complementarities. However, if it is impossible to bound these shares from below, then it also would be impossible to generically constrain the cannibalization on the cost side below the positive complementarities that arise from the demand side.

To derive a sufficient condition for SCD-T with Hicks neutral productivity, we consider how this marginal value changes in productivity z. In particular, one again define a function $\tilde{g}$ such that

$$g(\mathbf{\Theta}; z) \equiv \frac{\tilde{g}(\mathbf{\Theta})}{z}.$$

Then, the marginal value of location j can be rewritten

$$
\begin{aligned}
D_j\pi_i\left(\mathcal{L};z\right) &= \sum_n \int_0^1 \sum_k \xi_{kijn} \frac{\partial \pi_n^*}{\partial \Theta_{kin}} \left(\frac{\tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{z} \right) \mathrm{d}t - f_{ij} \\
&= \sum_n \int_0^1 \sum_k \xi_{kijn} \pi_n^{*\prime} \left(\frac{\tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{z} \right) \frac{\partial \tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{\partial \Theta_{kin}} \frac{1}{z} \mathrm{d}t - f_{ij}
\end{aligned}.
$$

Comparing this marginal value for two otherwise identical agents with productivities z_1 and z_2, where $z_1 \leq z_2$, it is sufficient for SCD-T to show that the marginal value of the location in j is higher for the agent with the higher productivity. We employ the same approach as the single-dimension case, now to each of the K partial derivatives of π_n^*.

$$
\begin{aligned}
& D_j\pi_i\left(\mathcal{L};z_2\right) - D_j\pi_i\left(\mathcal{L};z_1\right) \\
&= \int_{z_1}^{z_2} \frac{\partial}{\partial z} \sum_n \int_0^1 \sum_k \xi_{kijn} \pi_n^{*\prime} \left(\frac{\tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{z} \right) \frac{\partial \tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{\partial \Theta_{kin}} \frac{1}{z} \mathrm{d}t\mathrm{d}z \\
&= \int_{z_1}^{z_2} \sum_n \int_0^1 \sum_k \xi_{kijn} \frac{\partial \tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{\partial \Theta_{kin}} \left[- \frac{\pi_n^{*\prime}\left(\frac{\tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{z}\right)}{z^2} \right] \left[\varepsilon_{\pi_n^{*\prime}} - 1 \right] \mathrm{d}t\mathrm{d}z \\
&= \int_{z_1}^{z_2} \int_0^1 \sum_n \left[- \frac{\pi_n^{*\prime}\left(\frac{\tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{z}\right)}{z^2} \right] \left[\varepsilon_{\pi_n^{*\prime}} - 1 \right] \sum_k \xi_{kijn} \frac{\partial \tilde{g}\left(\boldsymbol{\xi}_{in}(j)t + \boldsymbol{\Theta}\left(\mathcal{L}\right)\right)}{\partial \Theta_{kin}} \mathrm{d}t\mathrm{d}z
\end{aligned}
$$

By the same argument, the term in the first set of square brackets is positive. Therefore, for this difference to be non-negative, it is sufficient for the term in the second set of square brackets to the positive. We thus derive the same sufficiency condition for SCD-T, namely that $\varepsilon_{\pi_n^{*\prime}} \geq 1$.

OA.4 The Branching Procedure

In this section, we provide the details about the basic branching procedure to maximize the objective function of a single agent and the generalized branching procedure to solve for the policy function mapping firm types to optimal strategies.

OA.4.1 The Basic Branching Procedure

At the heart of the branching procedure is a "branching step" applied to a CDCP for which the squeezing procedure has converged. The branching step takes an undetermined item ℓ such that $\ell \in \overline{\mathcal{L}}^{(K)} \setminus \underline{\mathcal{L}}^{(K)}$ and forms two subproblems, or "branches:" one in which ℓ is included in $\mathcal{L}^\star$ (i.e., added to $\underline{\mathcal{L}}$) and one in which it is excluded from $\mathcal{L}^\star$ (i.e., excluded from $\overline{\mathcal{L}}$).[19] The two fixed points resulting from applying the squeezing procedure to the bounding sets of each

[19]Any item ℓ such that $\ell \in \overline{\mathcal{L}}^{(K)} \setminus \underline{\mathcal{L}}^{(K)}$ can be chosen to initiate the branching procedure.

subproblem are the optimal decision sets *conditional* on the assumed inclusion or exclusion of ℓ. The optimal location set of the original CDCP is then the conditional optimal location set that yields the higher value of π.

In cases where the fixed point of at least one of the subproblems does not contain two identical sets, the branching step can be applied *recursively*. In particular, within each subproblem, we focus on another undetermined item ℓ' and create two sub-sub-problems. Recursively applying the branching step in such a way creates a "tree," where the terminal nodes are subproblems for which the squeezing procedure has converged to a bounding set pair where the lower bound and upper bound are equal.

We now formally define the branching step which uses the squeezing step from Definition 5:

Definition 8 (Branching step)**.** Given bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, select some element $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$. The mapping B^a is given by

$$B^a([\underline{\mathcal{L}}, \overline{\mathcal{L}}]) \equiv \left\{ S^{a(K)}\left([\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}]\right), S^{a(K)}\left([\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}]\right) \right\}$$

The mapping B^b is given by

$$B^b([\underline{\mathcal{L}}, \overline{\mathcal{L}}]) \equiv \left\{ S^{b(K)}\left([\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}]\right), S^{b(K)}\left([\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}]\right) \right\}$$

For given initial bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, we denote the operator of applying the branching step until global convergence by $B^{a(K)}([\underline{\mathcal{L}}, \overline{\mathcal{L}}])$ and $B^{b(K)}([\underline{\mathcal{L}}, \overline{\mathcal{L}}])$, respectively. Global convergence of the branching step occurs when the stopping condition $\underline{\mathcal{L}} = \overline{\mathcal{L}}$ is met on each branch.[20]

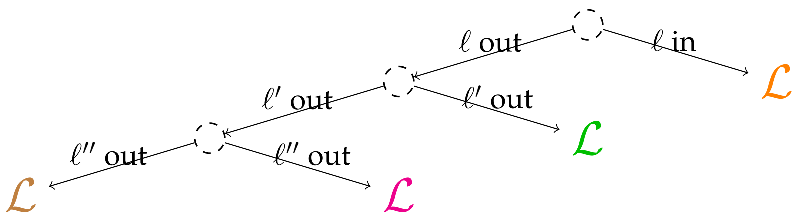
Suppose the return function exhibits SCD-C from above.[21] Given an initial bounding pair $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ with $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}$, the globally converged result $B^{a(K)}([\underline{\mathcal{L}}, \overline{\mathcal{L}}])$ is a collection of branch-specific optimal decision sets. The cardinality of the set is the number of branches. Among these conditionally optimal decision sets, the one yielding the highest value of π is the optimal location set solving the original CDCP. Note that contrary to the squeezing procedure, the branching procedure *always* identifies the optimal decision set.

For exposition, suppose the return function satisfies SCD-C, and consider Figure OA.3 which shows an example of a tree created by the branching procedure. It starts with a bounding set pair for which the squeezing procedure has converged, but there still remain undetermined items. One of these, ℓ, is selected. Two branches based on this item are formed. The left hand branch corresponds to the subproblem where ℓ is presumed to be excluded from the optimal decision set, while the right hand branch corresponds to the subproblem where ℓ is presumed

[20]Note that the definitions of the branching steps B^a and B^b suppose convergence of the squeezing procedure, so they are defined only when this convergence occurs. When then return function exhibits SCD-C, the squeezing procedure always converges.

[21]The same logic applies with $B^{b(K)}$ when the underlying return function exhibits SCD-C from below.

Figure OA.3: An Example Path of the Branching Procedure



Notes: The figure shows an example of a tree of subproblems resulting from applying the branching procedure recursively. Convergence on a single branch occurs when the squeezing procedure returns a conditionally optimal set, denoted by the colored $\mathcal{L}$s. The final output of the full recursive algorithm is the collection of all conditionally optimal sets.

to be included. The squeezing procedure is reapplied in each branch. On the right hand branch, the squeezing procedure delivers a bounding pair where $\underline{\mathcal{L}} = \overline{\mathcal{L}}$, yielding the orange $\mathcal{L}$. This location set is optimal conditional on the requirement that ℓ must be included. On the other hand, convergence of the squeezing procedure in the left hand branch does not deliver an optimal decision set. The returned bounding pair is still such that there are strictly more items in the upper bound than lower bound set. The branching procedure therefore branches again, this time selecting the still undetermined item ℓ' on which to branch.

Repeating the squeezing procedure on both branches, the right hand branch once again delivers a conditionally optimal decision set, the green $\mathcal{L}$. This green location set $\mathcal{L}$ is optimal conditional on both ℓ and ℓ' being included in the decision set. Again, the left hand branch does not deliver an optimal set, so the branching step is applied one last time, this time branching on item ℓ''. This branch yields conditionally optimal decision sets, the brown and pink $\mathcal{L}$s. The first is optimal conditional on ℓ, ℓ', and ℓ'' all being excluded. Likewise, the second is optimal conditional on excluding ℓ and ℓ', but including ℓ''. As a final step, all conditionally optimal sets must be manually compared, by evaluating the return function with each. The location set yielding the highest value is the global optimum.

To summarize the branching procedure, consider a CDCP as defined in equation (6) and let the bounding pair $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ be such that $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}$.

Then, if π exhibits SCD-C from above,

$$\mathcal{L}^\star = \underset{\mathcal{L} \in B^{a(K)}(S^a([\underline{\mathcal{L}}, \overline{\mathcal{L}}]))}{\arg\max} \pi(\mathcal{L}; \mathbf{z}) \, .$$

while if π exhibits SCD-C from below,

$$\mathcal{L}^\star = \underset{\mathcal{L} \in B^{b(K)}(S^b([\underline{\mathcal{L}}, \overline{\mathcal{L}}]))}{\arg\max} \pi(\mathcal{L}; \mathbf{z}) \, .$$

OA.4.2 The Generalized Branching Procedure

We define the generalized branching step which uses the generalized squeezing step in Definition 7 as follows:

Definition 9 (Generalized branching step)**.** Given a 4-tuple $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]$ and some $\ell \in M$, The mapping B^a is given by

$$B^a([(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]) \equiv S^{a(K)}([\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}, \emptyset, Z]) \cup S^{a(K)}([\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}, \emptyset, Z]) \,.$$

where $S^{a(K)}$ denotes recursively applying S^a until convergence.
The mapping B^b is given by

$$B^b([(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]) \equiv S^{b(K)}([\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}, \emptyset, Z]) \cup S^{b(K)}([\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}, \emptyset, Z]) \,.$$

where $S^{b(K)}$ denotes recursively applying S^b until convergence.

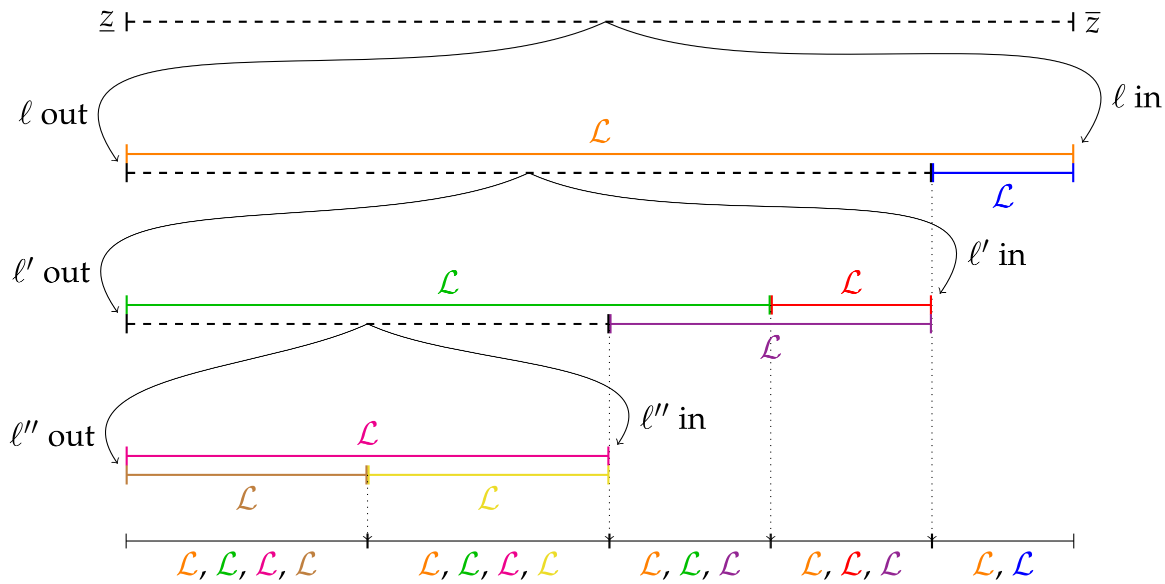
Given an initial 4-tuple $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]$ and an undetermined item $\ell \in M$, the generalized branching step creates two branches or subproblems. The first supposes that ℓ is included in the optimal decision set, while the second supposes that it is excluded. Note that, since the generalized squeezing step returns a set of 4-tuples, the generalized branching step combines them together with set union. To each of these two subproblems we apply the generalized squeezing procedure until global convergence obtaining a collection of 4-tuples which exhaustively partition the original region Z.[22] Each branch may now contain different partitions of the original type space.

On either branch, if there are any 4-tuples with undetermined items, the generalized branching step can be applied again. The generalized branching procedure consists in recursively applying the generalized branching step this way, where recursion stops on a given 4-tuple when the bounding sets for that 4-tuple coincide. Global convergence occurs when bounding sets coincide for 4-tuples on all branches. Then, the output of the generalized branching procedure is a collection of 4-tuples each of the form $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, \emptyset), Z]$.

For illustration, Figure OA.4 depicts the process of applying the generalized branching procedure to an initial 4-tuple $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]$. The initial 4-tuple specifies lower and upper bound sets $(\underline{\mathcal{L}}, \overline{\mathcal{L}})$ over the entire dotted interval Z between $\underline{z}$ and $\overline{z}$. Applying the generalized branching step once, the problem is divided into two subproblems: one corresponding to requiring item ℓ to be excluded, and the other requiring that item ℓ be included. In the subproblem on the right branch, the squeezing procedure identifies the single (orange) location set that is optimal

[22] The definitions of the branching steps B^a and B^b suppose convergence of the generalized squeezing procedure, and are therefore defined only when convergence occurs.

Figure OA.4: An Example Outcome of the Generalized Branching Procedure



Notes: The figure illustrates a recursive application of the generalized branching procedure. At each application, one undetermined item is selected on which to branch, yielding conditional policy functions. Each colored $\mathcal{L}$ represent a different decision set. Branching continues until no undetermined items remain for all types on all branches. Conditionally optimal decision sets for each type are ultimately gathered at the bottom of the figure.

for the whole interval conditional on including ℓ. On the left branch, ℓ is excluded. In this case, convergence from the squeezing procedure delivers a policy function only for the highest types $z \in Z$, identifying the (blue) optimal decision set. Undetermined elements remain for the lower types of the range. The branching step is thus reapplied for this subsection of the original interval, selecting a second undetermined element, ℓ'. This procedure repeats until no undetermined elements remain in any of the branches for any type $z \in Z$.

Now consider the entire initial region Z, repeated at the bottom of the graphic. The repeated application of the squeezing procedure to smaller and smaller subregions of the type space creates subregions of the overall type space that share several conditional optimal policy functions. We show the conditional optimal policy function that apply to each subregion. For each subregion, we now manually choose which of the associated conditional policy functions maximizes the return function for each type in the subregion. Piecing together the so chosen optimal policy functions for each interval yields the optimal policy function that solves the original CDCP on the interval $[\underline{z}, \overline{z}]$.

OA.5 The Pollak Demand System

Our quantitative application specializes the demand system of our general framework from Section 2 to the standard constant elasticity (CES) demand system. The CES demand system implies constant markups over marginal costs. In this section, we discuss the Pollak demand system which is also nested by our general formulation in Section 2, but implies variable markups over marginal costs instead.

The class of demand systems introduced by Pollak (1971) has become popular in the literature studying variable markups.[23] Consider the following demand function for a good ω from the Pollak class which is frequently used in quantitative applications (e.g., Arkolakis et al. (2019)) and first appeared in Klenow and Willis (2016):

$$q_n\left(p_n(\omega)\right) = Q_n D\left(p_n\left(\omega\right)/P_n\right) = \left(p_n\left(\omega\right)/P_n^{\star}\right)^{-\sigma} + \gamma, \quad \text{where} \quad Q_n = 1 \tag{OA.5}$$

where $\gamma < 0$. The demand in equation (OA.5) has a variety of appealing features. First, it features a choke price, $P_n^{\star}$, which implies that entry into each destination market n is guaranteed only for the firms with low enough marginal costs, $c_n \leq P_n^{\star}$. In other words, a firm does not necessarily serve all countries, but self-selects into export markets consistent with the data (see also Arkolakis et al. (2019)). Second, asymptotically, the elasticity of demand is constant, which allows the model to fit the Pareto-size tails of firm distribution for the largest firms and exporters, a key feature of the data (see Arkolakis (2016); Amiti et al. (2019)). Finally, under very general conditions, it implies that markups increase with firm size, a salient finding of recent investigations on the relationship of firm size and firm markups (see De Loecker et al. (2016)).

The pricing rule implied by the demand function in equation (OA.5) is given by

$$p_n\left(\omega\right) = \frac{\sigma}{\left(\sigma - 1\right) + \left(p_n\left(\omega\right)/P_n^{\star}\right)^{\sigma}} c,$$

where the markup is decreasing in the firm's marginal cost, c, in contrast with the CES demand system used in our quantitative application. At one extreme, when the firm's marginal cost is equal to the choke price, it sets a price at marginal cost. Firms with even higher marginal cost do not participate in the market. At the other extreme, very productive firms with marginal cost approaching zero have markings approaching $\frac{\sigma}{\sigma-1}$.Thus, the key parameter is σ, which parsimoniously captures the relationship between firm productivity, size, and markup.

Finally, the statistic necessary to establish whether SCD-C holds for a given CDCP (cf. Section

[23] See, for example, Simonovska (2015), Arkolakis et al. (2019), and Behrens et al. (2020).

OA.3.2) takes the convenient form:

$$\varepsilon_q(p_n(\omega))\frac{\mathrm{d}\ln p_n(\omega)}{\mathrm{d}\ln c_{in}(\omega)} = \underbrace{\frac{\sigma}{1-(p_n(\omega)/P_n^\star)^\sigma}}_{\text{Demand Elasticity}}\underbrace{\frac{(\sigma-1)+(p_n(\omega)/P_n^\star)^\sigma}{(\sigma-1)+(\sigma+1)\,(p_n(\omega)/P_n^\star)^\sigma}}_{\text{Passthrough}} \geq \sigma.$$

The lower bound σ makes it straightforward verify if the firm's problem satisfies both SCD-C from below and SCD-T. In particular, in our quantitative application, as long as σ is large enough to exceed $1+\theta$, the firm problem exhibits both SCD-C from below and SCD-T, both with the Pollak and CES demand systems.

OA.6 CDCPs in the Literature

In this section, we present two CDCPs from the literature and show that they satisfy our SCD-C condition. First, we introduce the simple plant location problem, a canonical, NP-hard problem in the operations research literature. Second, we show that a version of the firm problem in Arkolakis et al. (2018) with fixed costs is a CDCP that satisfies SCD-C.

OA.6.1 The Simple Plant Location Problem

The Simple Plant Location Problem ("SPLP") or Un-capacitated Plant Location Problem ("UPLP") refers to a general class of operation research problems concerned with the optimal choice of production locations to minimize transportation costs while serving a set of spatially distributed demand points. The SPLP is a canonical problem in Operations Research (e.g., p-Center and p-Median problems) (see Krarup and Pruzan (1983), Cornuéjols et al. (1983), Owen and Daskin (1998), and Verter (2011) for surveys). While the problem has been shown to be NP-hard, there are many heuristic solution methods.

In this section, we show that the SPLP can be expressed as a CDCP as in Definition 1 that satisfies the SCD-C property and can hence be solved using our methods.

We outline the setup of the SPLP as outlined in Balinski (1965). Consider an economy with L of potential facility locations and a set N demand points. Opening a production facility in location ℓ incurs a fixed cost $f_\ell \geq 0$. The marginal cost of serving destination n from the facility in some location ℓ is $c_{\ell n} \geq 0$. Each location demands the same fixed quantity of the good and all locations need to be served. The Boolean choice variable is $\lambda_{\ell n}$ which is 1 if market n is served from location ℓ and zero otherwise. The SPLP is formulated as the problem of minimizing the total cost of serving all demand points:

$$\min_{\{\lambda_{\ell n}\}_{\ell\in L, n\in N}} \sum_{\ell\in L}\sum_{n\in N} c_{\ell n}\lambda_{\ell n} + \sum_{\ell\in L} f_\ell\theta_\ell \tag{OA.6}$$

$$\text{subject to} \sum_{\ell \in L} \lambda_{\ell n} = 1 \ \forall n$$
$$\theta_\ell \geq \lambda_{\ell n} \ \forall \ell, n$$
$$\lambda_{\ell n}, \theta_\ell \in \{0, 1\} \ \forall \ell, n$$

The equality constraint is imposed to ensure each market is served by exactly one production site; the inequality constraint ensures that the relevant fixed costs are paid for every production location in operation.

We rewrite the above SPLP to show that it fits our definition of a CDCP. First, we define $c'_{\ell n} = \max_\ell f_\ell + \max_{\ell n} c_{\ell n} - c_{\ell n}$. We denote by $\mathcal{L}$ the set of production locations, so that $\mathcal{L} \subseteq L$.For a given production location set $\mathcal{L}$ the profit from a producing in location ℓ is then given by:

$$\pi_\ell(\mathcal{L}) = \mathbb{1}\left(\ell \in \mathcal{L}\right) \left[\sum_{n \in N} c'_{\ell n} \mathbb{1}\left(\max_{k \in \mathcal{L}} c'_{kn} \leq c'_{\ell n}\right) - f_\ell\right] \tag{OA.7}$$

The firms overall profit is the sum of the profits of all its production locations, i.e., $\pi(\mathcal{L}) = \sum_{\ell \in \mathcal{L}} \pi_\ell(\mathcal{L})$. The problem in equation (OA.6) can then be written as

$$\mathcal{L}^\star = \arg\max_{\mathcal{L} \in L} \pi(\mathcal{L}),$$

which corresponds to Definition 1. Note that the definition of $c'_{\ell n}$ above ensures that at least one production location always operates and that every demand point is always served.

Next, we show that the firm's objective satisfies the SCD-C property. In fact, it satisfies the stronger "submodularity" property which is sufficient for SCD-C from above. From the perspective of the firm, adding an additional production location always weakly decrease the number of demand points served by any previously existing production location. As a result, the location-specific profit in equation (OA.7) is weakly decreasing in the total number of production locations. To see this formally, consider the following Lemma:

Lemma 1. *The profit function in the simple plant location problem is submodular.*

Proof. Consider two sets $\emptyset \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subseteq L$. But then notice that for some facility location $\ell \in \mathcal{L}_2$:

$$\begin{aligned}
\pi_\ell(\mathcal{L}_1) &= \mathbb{1}\left(\ell \in \mathcal{L}_1\right) \left[\sum_{n \in N} c'_{\ell n} \mathbb{1}\left(\max_{k \in \mathcal{L}_1} c'_{kn} \leq c'_{\ell n}\right) - f_l\right] \\
&\leq \mathbb{1}\left(i \in \mathcal{L}_2\right) \left[\sum_{n \in N} c'_{\ell n} \mathbb{1}\left(\max_{k \in \mathcal{L}_2} c'_{kn} \leq c'_{\ell n}\right) - f_\ell\right] = \pi_\ell(\mathcal{L}_2)
\end{aligned}$$

where the inequality holds since $\max_{k \in \mathcal{L}_2} c'_{kn} \leq \max_{k \in \mathcal{L}_1} c'_{kn}$. Now consider two different sets,

$\mathcal{L}_1' = \mathcal{L}_1 \backslash k$ and $\mathcal{L}_2' = \mathcal{L}_2 \backslash k$ where $k \in \mathcal{L}_2$. But then

$$\pi_\ell(\mathcal{L}_1') \leq \pi_\ell(\mathcal{L}_2')$$

But then since inequalities are closed under addition:

$$\pi_\ell(\mathcal{L}_1) - \pi_\ell(\mathcal{L}_1') \leq \pi_\ell(\mathcal{L}_2) - \pi_\ell(\mathcal{L}_2')$$

so that the profit of the facility in location ℓ, π_ℓ, is submodular on the set L. Since the submodularity property is closed under addition, the overall profit function, $\pi(\mathcal{L}) = \sum_{\ell \in \mathcal{L}} \pi_\ell$, is submodular on the set L. $\square$

OA.6.2 The Firm Problem in Arkolakis et al. (2018)

For a given set of markets N, the firm problem that appears in Arkolakis et al. (2018) is combinatorial since the marginal value of each production location depends on the other available production locations. However, it is a trivial problem since there are no fixed costs. Accordingly, the marginal value of each production location is never negative under any circumstances, so firms operate in all locations.

If fixed costs are incorporated in the firm decision, so that the firm headquartered in i must pay a fixed cost $f_{i\ell}$ to operate a production location in ℓ, expected profit function is then

$$\pi_i\left(\mathcal{L}, \mathbf{z}\right) = \sum_{n \in N} X_n \left(\frac{\sigma}{\sigma - 1} c_{in}\left(\mathcal{L}, \mathbf{z}\right)\right)^{1-\sigma} - \sum_{\ell \in \mathcal{L}} f_{i\ell} \qquad , \qquad c_{in}\left(\mathcal{L}, \mathbf{z}\right) = \min_{\ell \in \mathcal{L}} \frac{\gamma_{i\ell} w_\ell d_{\ell n}}{z_\ell}.$$

Accordingly, the marginal value of a production location ℓ is

$$D_\ell \pi_i\left(\mathcal{L}, \mathbf{z}\right) = \sum_{n \in N} \left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} \left[c_{in}\left(\mathcal{L} \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} - c_{in}\left(\mathcal{L} \setminus \{\ell\}, \mathbf{z}\right)^{1-\sigma}\right] - f_{i\ell}.$$

To establish monotone substitutes, we compare this marginal value for two arbitrary production location sets $\mathcal{L}_1 \subseteq \mathcal{L}_2$, and argue that the marginal benefit to $\mathcal{L}_1$ must exceed the marginal benefit to $\mathcal{L}_2$. We begin by note that, for each destination market n,

$$0 < c_{in}\left(\mathcal{L} \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} - c_{in}\left(\mathcal{L} \setminus \{\ell\}, \mathbf{z}\right)^{1-\sigma} \quad \Leftrightarrow \quad \ell = \arg \min_{\ell \in \mathcal{L} \cup \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell n}}{z_\ell} \quad .$$

Then, we can identify the locations $\tilde{n}$ for which ℓ is the least-cost location as

$$\tilde{N}_2 \equiv \left\{\tilde{n} \in N \mid \ell = \arg \min_{\ell \in \mathcal{L}_2 \cup \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell \tilde{n}}}{z_\ell}\right\} \quad .$$

Note that, for any $\tilde{n} \in \tilde{N}$,

$$c_{i\tilde{n}}\left(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} - c_{i\tilde{n}}\left(\mathcal{L}_2 \setminus \{\ell\}, \mathbf{z}\right)^{1-\sigma} \geq 0.$$

Considering this difference for $\mathcal{L}_1$, note that ℓ must also be the least-cost supplier, since $\mathcal{L}_1 \subseteq \mathcal{L}_2$. Thus,

$$c_{i\tilde{n}}\left(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} = c_{i\tilde{n}}\left(\mathcal{L}_1 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma}.$$

On the other hand,

$$\begin{aligned} \min_{\ell \in \mathcal{L}_1 \setminus \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell\tilde{n}}}{z_\ell} &\geq \min_{\ell \in \mathcal{L}_2 \setminus \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell\tilde{n}}}{z_\ell} \\ \Rightarrow c_{i\tilde{n}}\left(\mathcal{L}_1 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} &\leq c_{i\tilde{n}}\left(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} \end{aligned}$$

since the best supplier in $\mathcal{L}_2 \setminus \{\ell\}$ may not be included in $\mathcal{L}_1 \setminus \{\ell\}$. Put together,

$$c_{i\tilde{n}}\left(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} - c_{i\tilde{n}}\left(\mathcal{L}_2 \setminus \{\ell\}, \mathbf{z}\right)^{1-\sigma} \leq c_{i\tilde{n}}\left(\mathcal{L}_1 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} - c_{i\tilde{n}}\left(\mathcal{L}_1 \setminus \{\ell\}, \mathbf{z}\right)^{1-\sigma}.$$

Next, consider any $n' \in \tilde{N}_2^c$, and note that ℓ is not the least cost-supplier. Then,

$$\begin{aligned} \ell' &\equiv \arg\min_{\ell \in \mathcal{L}_2 \cup \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell n'}}{z_\ell} \neq \ell \\ \Rightarrow \ell' &= \arg\min_{\ell \in \mathcal{L}_2 \setminus \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell n'}}{z_\ell} \\ \Rightarrow 0 &= c_{in'}\left(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} - c_{in'}\left(\mathcal{L}_2 \setminus \{\ell\}, \mathbf{z}\right)^{1-\sigma} \end{aligned}.$$

Intuitively, adding ℓ to $\mathcal{L}_2$ does not provide the firm any benefit for market n', since it is not the least-cost supplier. However, the firm can always do weakly better in market n' if ℓ is added to $\mathcal{L}_1 \setminus \{\ell\}$. In other words,

$$\begin{aligned} \min_{\ell \in \mathcal{L}_1 \setminus \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell n'}}{z_\ell} &\geq \min_{\ell \in \mathcal{L}_1 \cup \{\ell\}} \frac{\gamma_{i\ell} w_\ell d_{\ell n'}}{z_\ell} \\ \Rightarrow c_{i\tilde{n}}\left(\mathcal{L}_1 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} &\leq c_{i\tilde{n}}\left(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z}\right)^{1-\sigma} \end{aligned}.$$

Putting everything together,

$$
\begin{aligned}
D_\ell \pi_i(\mathcal{L}_2, \mathbf{z}) &= \sum_{n \in N} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left[c_{in}(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z})^{1-\sigma} - c_{in}(\mathcal{L}_2 \setminus \{\ell\}, \mathbf{z})^{1-\sigma}\right] - f_{i\ell} \\
&= \sum_{\tilde{n} \in \tilde{N}_2} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left[c_{i\tilde{n}}(\mathcal{L}_2 \cup \{\ell\}, \mathbf{z})^{1-\sigma} - c_{i\tilde{n}}(\mathcal{L}_2 \setminus \{\ell\}, \mathbf{z})^{1-\sigma}\right] - f_{i\ell} \\
&\leq \sum_{\tilde{n} \in \tilde{N}_2} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left[c_{i\tilde{n}}(\mathcal{L}_1 \cup \{\ell\}, \mathbf{z})^{1-\sigma} - c_{i\tilde{n}}(\mathcal{L}_1 \setminus \{\ell\}, \mathbf{z})^{1-\sigma}\right] - f_{i\ell} \\
&\leq \sum_{n \in N} \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left[c_{in}(\mathcal{L}_1 \cup \{\ell\}, \mathbf{z})^{1-\sigma} - c_{in}(\mathcal{L}_1 \setminus \{\ell\}, \mathbf{z})^{1-\sigma}\right] - f_{i\ell} = D_\ell \pi_i(\mathcal{L}_1, \mathbf{z})
\end{aligned}
$$

Thus, the firm's profit function displays monotone substitutes, and thus SCD-C from below.

OA.7 Additional Technical Discussion

In this section, we offer a technical discussion of our squeezing algorithm and compare it directly to the procedure outlined in Jia (2008). We first rewrite our algorithm using lattice algebra to make it directly comparable to the method in Jia (2008).

OA.7.1 Comparison with Jia (2008)

Jia (2008) introduced a "reduction method" for supermodular CDCPs to reduce the set of potentially optimal strategies iteratively. Jia (2008) outlines this method using lattice algebra. In this section, for direct comparison with Jia (2008), we first translate our definition of CDCPs which uses power sets of discrete sets into lattice notation and then demonstrate our contribution relative to Jia (2008). Since Jia (2008) only discusses the CDCP of a single agent, we abstract from type heterogeneity throughout this section.

CDCPs on a Lattice Consider a CDCP as in Definition 1: an agent maximizes the return function $\pi(\cdot)$ by choosing $\mathcal{L} \in \mathscr{P}(L)$ where L is a discrete set. We denote the return-function-maximizing set by $\mathcal{L}^\star$, so that $\mathcal{L}^\star = \arg\max_{\mathcal{L} \in \mathscr{P}(L)} \pi(\mathcal{L})$.

We construct an isomorphic problem in which the agent chooses elements from the Boolean lattice to maximize the same return function. First, we assign a unique index number $i = 1, \ldots, N$ to every element in the discrete set L. Second, we construct a function $g : \mathcal{B}^N \to \mathscr{P}(L)$ that maps from the Boolean lattice of dimension N, $\mathcal{B}^N$, to the power set of $\mathscr{P}(L)$. In particular, g assigns to any element $\mathcal{L} \in \mathscr{P}(L)$ the Boolean vector I such that $I_i = 1$ if the element with index number i is contained in $\mathcal{L}$ and $I_i = 0$ otherwise.

Third, we define a transformed return function $\tilde{\pi} : \mathcal{B}^N \mapsto \mathbb{R}$, s.t. $\tilde{\pi}(\cdot) = \pi(g(\cdot))$. The problem of finding the element $I^\star = \arg\max_{I \in \mathcal{B}^N} \tilde{\pi}(I)$ of the Boolean lattice is identical to the problem of finding $\mathcal{L}^\star = \arg\max_{\mathcal{L} \in \mathscr{P}(L)} \tilde{\pi}(\mathcal{L})$ and $g(I^\star) = \mathcal{L}^\star$ holds by construction. Importantly, the set of all Boolean vectors in $\mathcal{B}^N$ along with the "pairwise comparison" norm, whereby $I \leq I'$ if and only if $I_\ell \leq I'_\ell$ $\forall \ell$, forms a *complete lattice* denoted by $U = \langle \mathcal{B}^N, \leq \rangle$.[24] Also note that a vector $I^\star$ always exists since $\mathcal{B}^N$ is a finite and complete lattice which always contains (at least) one vector for which $\tilde{\pi}(\cdot)$ obtains its largest value over $\mathcal{B}^N$.

Jia (2008) presented an iterative algorithm that narrows the set of potential solution vector $I \in \mathcal{B}^N$ of the CDCP as long as $\tilde{\pi}(\cdot)$ exhibits supermodularity. We now introduce the iterative mapping at the heart of the approach in Jia (2008).

Constructing a Function with a Guaranteed Fixed Point Consider a Boolean vector $I \in \mathcal{B}^N$.Denote $I = \{I_1, \ldots, I_N\}$ where I_i is the value of the nth entry of the vector I.Using this notation, the optimality of $I^\star$ implies that:

$$\tilde{\pi}(\{I_1^\star, ..., I_i^\star, ..., I_n^\star\}) \geq \tilde{\pi}(\{I_1^\star, ..., I_i, ..., I_n^\star\}) \quad \forall i. \tag{OA.8}$$

Introducing a Boolean version of the marginal value operator (cf. Definition 2), equation (OA.8) implies the following for the optimal vector $I^\star$:

$$I_i^\star = \begin{cases} 1 & \text{if } D_i\tilde{\pi} \geq 0 \\ 0 & \text{if } D_i\tilde{\pi} < 0 \end{cases}.$$

Now define the following mapping $\Lambda : \mathcal{B}^N \mapsto \mathcal{B}^N$ such that

$$I^{k+1} = \Lambda(I^k) = \{\Lambda_1(I^k), \ldots, \Lambda_N(I^k)\} \quad \text{where} \quad \Lambda_i(I^k) = I_i^{k+1} = 1[D_i\pi(I^k) \geq 0], \tag{OA.9}$$

and k indexes the number of times the function has been applied to some initializing vector I_0. Importantly, by construction, the return-function maximizing vector $I^\star$ is a fixed point of Λ, so that $I^\star = \Lambda(I^\star)$. However, the mapping Λ may have (many) other fixed points in addition to $I^\star$. Next, we show the role of the supermodularity assumption in Jia (2008) in allowing a characterization of the set of fixed points of Λ.

Jia (2008) and the Role of Supermodularity The assumption in Jia (2008) that gives rise to the CDCP reduction method is that the return function $\tilde{\pi}(\cdot)$ exhibits the supermodularity property. Under the supermodularity assumption, for two ordered Boolean vectors $I \leq I' \in \mathcal{B}^N$,

$$D_i\tilde{\pi}(I) \leq D_i\tilde{\pi}(I') \quad \forall i = 1, \ldots, N.$$

[24] See Kusraev and Kutateladze (2012) for a discussion of general results in Boolean valued analysis.

As a result, the function Λ is increasing over the partially ordered set $\mathcal{B}^N$ (alternative names for such functions are order-preserving, isotone, and monotone functions). In particular for any $I \leq I' \in \mathcal{B}^n$ it is easy to see that $\Lambda(I) \leq \Lambda(I') \in \mathcal{B}^n$, so the order of the two vectors is preserved by the application of the map. The "increasing" property of the mapping Λ implied by the supermodularity assumption allows Jia (2008) to invoke the main theorem of Tarski (1955) which we restate here for convenience:

Theorem 3. *Let*
(1) $U = \langle A, \leq \rangle$ be a complete lattice.
(2) f be an increasing function from A to A
(3) P be the set of all fixed points of f
Then the set P is not empty and the system $\langle P, \leq \rangle$ is a complete lattice.

Tarski's theorem implies that the function Λ has at least one fixed point, and that the set of *all* its fixed points forms a complete lattice.

In the context of our CDCP, the importance of Tarski's theorem is not that it proves the existence of a fixed point (since $I^\star$ is a fixed point of the mapping Λ by construction), but that it shows that the *set* of all fixed points of Λ, P, forms a *complete* lattice. The complete lattice property of the set of fixed points P implies that the set P has an upper and lower bound, $I^{UB} := \sup P$ and $I^{LB} := \inf P$ which bound the optimal vector, i.e.,$I^{LB} \leq I^\star \leq I^{UB}$. The Jia (2008) reduction method identifies the set P of fixed points by identifying its least and its greatest element. Finding the return-function-maximizing vector in P is usually easier than finding it in $\mathcal{B}^N$ since P is a subset of $\mathcal{B}^N$. We briefly describe how to find the least and greatest element of P. By definition, for any $I \in P$:

$$\Lambda(I) = I.$$

Consider the least fixed point in $I^{LB} \in P$. Note that for all $I \in P$, $I^{LB} \leq I$ by definition. Now consider the vector of all zeros $I_0 = [0, \dots, 0] = \inf \mathcal{B}^N$. and note:

$$I_0 \leq I^{LB} \leq I^\star.$$

Since the mapping Λ is increasing:

$$I^1 = \Lambda(I_0) \leq \Lambda(I^{LB}) = I^{LB}.$$

Applying the mapping Λ $K \leq N$ times to the vector I_0 eventually yields I^{LB}.

$$I^K = \Lambda(I^{K-1}) = I^{LB}.$$

The insight in Jia (2008) is that as long as the mapping Λ is increasing, starting with the least element in $\mathcal{B}^N$ and iterating is guaranteed to identify I^{LB} since applying an increasing mapping to two ordered vectors always preserves their order. At each iteration, at least one additional entry in I^k is set to 1 and, since Λ is increasing, it never reverts back to 0 in any subsequent application of Λ. As a corollary, finding I^{LB}never takes more than N iterations. In a similar way, iteratively applying Λ to $I_1 = \sup \mathcal{B}^N = [1, \ldots, 1]$ identifies I^{UB}.

The first contribution of our paper is to show that the above argument also applies to return functions $\tilde{\pi}$ that satisfy single crossing differences from below, a condition that is weaker than the supermodularity assumption imposed in Jia (2008). A function $\tilde{\pi}$ satisfies single crossing differences from below, if for all i and decision sets $I \leq I' \in \mathcal{B}^N$,

$$D_i \pi (I) \geq 0 \Rightarrow D_i \pi \left(I'\right) \geq 0.$$

The SCD-C from below property is necessary and sufficient for the mapping Λ to be increasing, whereas the supermodularity property is merely sufficient. As long as the mapping Λ remains increasing, the above argument for identifying I^{LB} and I^{UB} remains valid.

A Submodular Return Function Now suppose that the return function obeys single crossing differences in choices from above (or negative complementarities) such that, for all i and decision sets $I \leq I' \in \mathcal{B}^n$,

$$D_i \pi \left(I'\right) \geq 0 \Rightarrow D_i \pi (I) \geq 0.$$

If the return function satisfies SCD-C form above, the mapping Λ defined in equation (OA.9) is no longer order-preserving (increasing) but order-reversing (decreasing) and the Tarski (1955) theorem no longer applies. This insight has stalled progress on solving problems with negative complementarities in the economics literature on multinational production despite their prevalence.

The central result in our paper is to demonstrate that the iterative application of Λ still identifies the greatest and least element of a set $P \subseteq \mathcal{B}^N$ such that $I^\star \in P$ when the underlying return function satisfies SCD-C from above.

To show this, we require a generalization of the concept of the fixed point to that of the *fixed edge* (see Klimeš (1981)):

Definition 10. Let f be a mapping of a partially ordered set U into itself and let $x \leq y$ be the elements of U. An ordered pair (x, y) is called a fixed edge of f if $f(x) = y$ and $f(y) = x$.

Intuitively, a fixed edge of a mapping are two points between which the mapping alternates. All fixed points are (degenerate) fixed edges.[25] Theorem 9 in Klimeš (1981) states the following Tarski-like theorem for decreasing functions:

[25]For increasing functions, all fixed edges are also fixed points.

Theorem 4. *Let*

(1) $U = \langle A, \leq \rangle$ be a complete lattice.

(2) f be a weakly decreasing function from A to A

(3) P be the set of all fixed edges of f

Then the set P is not empty and the system $\langle P, \leq \rangle$ is a complete lattice.

The theorem establishes that any decreasing mapping has a set of fixed edges that is itself a complete lattice. The set of fixed points of f is a subset of the set of fixed edges of f. Importantly, Theorem 4 shows that there exists a maximal element $I^{UB}$and a minimal element I^{LB} in P such that $(I^{UB}, f(I^{UB}))$ and $(I^{LB}, f(I^{LB}))$ are fixed edges of the function f.

To build intuition for the Klimeš (1981) theorem, consider the mapping defined in equation (OA.9) above, applied to a profit function π that satisfies SCD-C from above. It is straightforward to show that Λ is a decreasing mapping in this case. We denote the set of its fixed edges by P_Λ. Define an auxiliary function $f = \Lambda(\Lambda(\cdot))$ and denote the set of fixed points of f by P_f. It is easy to show that f is an *increasing* function on the same domain. As a result, the theorem in Tarski (1955) applies to f so that P_f is a non-empty and complete lattice.

First, note that trivially any fixed point of Λ is also a fixed point of g since:

$$I = \Lambda(I) \Rightarrow I = \Lambda(\Lambda(I)) = f(I),$$

so that the set of fixed points of Λ is a subset of the set of fixed points of f, i.e., $P_\Lambda \subseteq P_f$. Conversely, for any fixed point I of f, we know that

$$I = f(I) = \Lambda(\Lambda(I)) := \Lambda(I').$$

But then (I, I') form a fixed edge of Λ.So all elements in P_f are fixed edges of Λ, in other words $P_\Lambda = P_f$. Consequently, applying the same algorithm as in Jia (2008) identifies the set of fixed points P_f of the auxiliary mapping. If P_f is a singleton, we have identified $I^\star$; if $P_f \subset \mathcal{B}^N$, we have reduced the number of potential solutions to the CDCP.

Our squeezing step in the main body of the paper leverages these insights to create a mapping of the power set of a discrete set into itself that is always increasing. In the next section, we present a lattice-based formulation of our squeezing step.

OA.7.2 Lattice Formulation of Main Result

We present a version of the squeezing step in Definition 5 that uses lattice algebra:

Definition. Consider a sublattice of the Boolean lattice $\mathcal{I} \subseteq \mathcal{B}^N$ and the set of all its sublattices denoted by $S(\mathcal{I})$. Then we define the following two sets:

$$\bar{\Gamma}_1(\mathcal{I}) = \{i : D_i\pi(\sup \mathcal{I}) < 0\} \quad \text{and} \quad \underline{\Gamma}_1(\mathcal{I}) = \{i : D_i\pi(\inf \mathcal{I}) \geq 0\}$$

Then, we define the mapping $\mathrm{E}_1 : S(\mathcal{I}) \to S(\mathcal{I})$ as

$$\mathrm{E}_1(\mathcal{I}) = \{I \in \mathcal{I} : I_i = 0 \text{ and } I_j = 1, \forall i \in \bar{\Gamma}_1(\mathcal{I}), \forall j \in \underline{\Gamma}_1(\mathcal{I})\}.$$

Similarly, we define the following two sets:

$$\bar{\Gamma}_2(\mathcal{I}) = \{i : D_i\pi(\sup \mathcal{I}) > 0\} \quad \text{and} \quad \underline{\Gamma}_2(\mathcal{I}) = \{i : D_i\pi(\inf \mathcal{I}) \leq 0\}$$

And the mapping $\mathrm{E}_2 : S(\mathcal{I}) \to S(\mathcal{I})$ as

$$\mathrm{E}_2(\mathcal{I}) = \{I \in \mathcal{I} : I_i = 1 \text{ and } I_j = 0, \forall i \in \bar{\Gamma}_2(\mathcal{I}), \forall j \in \underline{\Gamma}_2(\mathcal{I})\}.$$

The mapping E_1 is an increasing mapping if $\tilde{\pi}$ exhibits SCD-C from below (or is supermodular.) Likewise, E_2 is an increasing mapping if $\tilde{\pi}$ exhibits SCD-C from above (or is submodular). In these cases, the main theorem in Tarski (1955) (cf. Theorem 3) then applies to both E_1 and E_2. Iterating on E_1 or E_2 converges to a set of vectors $\mathcal{I}^K$ which form a complete lattice. We can define $I^{LB} = \inf \mathcal{I}^K$ and $I^{UB} = \sup \mathcal{I}^K$, which are vectors such that $I^{LB} \subseteq P \subseteq I^{UB}$. Note that different from the function Λ defined in Jia (2008), the two mappings defined in the squeezing step are always increasing as long as the return function satisfies the appropriate version of SCD-C.

OA.7.3 Determinants of the Effectiveness of the Algorithm

In Section (5.4), we provided computational results for the convergence speed of the squeezing and branching algorithms. We showed that computational time generally increases as we moved from positive to negative complementarities. The lattice-based analysis in Section (OA.7.1) provides some intuition for why this is the case: with negative complementarities (SCD-C from above), the set of potential solutions contains not just fixed points but also fixed edges and hence tends to be larger. In this section, we use a simple example CDCP to illustrate the role of the distribution of location-specific payoffs ("geography") and the type of complementarity for the effectiveness of the algorithm. Instead of applying the squeezing step, we instead apply the mapping in equation (OA.9) since it is easier to write out its individual steps. However, the intuition behind the convergence of the squeezing step and the mapping in equation (OA.9) is the same.

As an example objective function, we use a simplified version of the objective function in Jia (2008). Consider a firm that maximizes a return function π by choosing a Boolean vector of dimension N, $I \in \mathcal{B}^N$, and the profit associated with each decision vector is given by the following function:

$$\pi(I) = \sum_{i=1,\dots,N} \pi_i \quad \text{where} \quad \pi_i = I_i \left(A_i + \gamma \sum_{k \neq i} I_k \right), \tag{OA.10}$$

where π denotes the agent's total return, π_i denotes the total return associated with entry i of the decision vector, the term A_i denotes the exogenous part of the return associated with entry i of the decision vector, and γ parameterizes the effect of other entries $k \neq i$ on the return associated with entry i. Notice that if $I_i = 0$ then $\pi_i = 0$ always. We refer to each i as a "location" so that the return function in equation (OA.10) could be that of a firm choosing a set of production locations.

There are two parameters that determine the shape of $\pi(\cdot)$: γ and $\{A_i\}_i$. The parameter γ determines whether the return function exhibits SCD-C from above or below. If $\gamma > 0$, the firm's objective function $\pi(\cdot)$ exhibits SCD-C from below or positive complementarities among locations; if the objective function exhibits SCD-C from above. The vector of "exogenous" returns $\{A_i\}_i$ determines the return to each location in the absence of any complementarities among locations.

We consider four different parameterizations of the return function in equation (OA.10) to illustrate the role of the two types of complementarities and the differences between two different types of stylized "geographies," or structures of payoffs across locations that we refer to as "cities" versus "countries." In the city geography, there is a hierarchy of locations in terms of A_i: a "central" location with a maximum A_i and then more and more remote regions with increasingly lower values of A_i mimicking, e.g., agglomeration economies that decay with distance from an urban center.The country geography setup features several locations with the same exogenous return A_i reflecting that there are several countries with similar productivities, e.g., the France and Germany may have a similar A_i while Chile and Argentina have A_i terms that are similar to one another but different from those of France and Germany. For both the city and country setups, we consider the case of SCD-C from below and above separately.

For simplicity, we assume a total of only six distinct locations $i = 1, ..., 6$ and show the explicit steps of obtaining a lower bound on the set of fixed fixed edges by iteratively applying the mapping in equation (OA.9).

Setup #1: City Geography and SCD-C From Below We set the exogenous return in equation (OA.10) equal to $A_i = i - 3$ and assume the return function exhibits SCD-C from below with $\gamma = 1$.But then applying the mapping from equation (OA.9) three times to the vector of all zeros, I_0, identifies the least element of the set of fixed points of the mapping:

$$I^1 := \Lambda(I_0) = [0, 0, 1, 1, 1, 1]$$

$$I^2 := \Lambda(I^1) = [1,1,1,1,1,1]$$
$$I^{LB} = I^3 := \Lambda(I^2) = [1,1,1,1,1,1] = I^2$$

so that the fixed point is found after three iterations. Starting from the vector of all 1s, I_1 :

$$I^{UB} = I^1 := \Lambda(I_1) = [1,1,1,1,1,1]$$

So that $I^{LB} = I^{UB} = I^\star$.

Setup #2: City Geography and SCD-C From Above We set the exogenous return in equation (OA.10) equal to $A_i = i - 3$ and assume the return function exhibits SCD-C from above with $\gamma = -1$. But then applying the mapping from equation (OA.9) three times to the vector of all zeros, I_0, identifies the least element of the set of fixed points of the mapping:

$$I^1 := \Lambda(I_0) = [0,0,1,1,1,1]$$
$$I^2 := \Lambda(I^1) = [0,0,0,0,0,1]$$
$$I^3 := \Lambda(I^2) = [0,0,0,1,1,1]$$
$$I^4 := \Lambda(I^3) = [0,0,0,0,1,1]$$
$$I^{LB} = I^5 := \Lambda(I^4) = [0,0,0,0,1,1] = I^4$$

so that the fixed point is found after five iterations.Starting from the vector of all 1s, I_1 :

$$I^1 := \Lambda(I_0) = [0,0,0,0,0,0]$$
$$I^2 := \Lambda(I^1) = [0,0,1,1,1,1]$$
$$I^3 := \Lambda(I^2) = [0,0,0,0,0,1]$$
$$I^4 := \Lambda(I^3) = [0,0,0,1,1,1]$$
$$I^5 := \Lambda(I^4) = [0,0,0,0,1,1]$$
$$I^{UB} = I^6 := \Lambda(I^5) = [0,0,0,0,1,1] = I^5$$

So that $I^{LB} = I^{UB} = I^\star$.

Setup #3: Country Geography and SCD-C From Below We set the exogenous return in equation (OA.10) equal to $A_i = -4\,\forall\, i \in [1,2]$ and $A_i = -2\,\forall\, i \in [3,4]$ and $A_i = 0\,\forall\, i \in [5,6]$. We also assume the return function exhibits SCD-C from below with $\gamma = 1$. But then applying the mapping from equation (OA.9) three times to the vector of all zeros, I_0, identifies the least element of the set of fixed points of the mapping:

$$I^1 := \Lambda(I_0) = [0,0,0,0,1,1]$$

$$
\begin{aligned}
I^2 &:= \Lambda(I^1) = [0,0,1,1,1,1] \\
I^3 &:= \Lambda(I^2) = [1,1,1,1,1,1] = I^2 \\
I^{LB} = I^4 &:= \Lambda(I^3) = [1,1,1,1,1,1] = I^3
\end{aligned}
$$

so that the fixed point is found after three iterations. Starting from the vector of all 1s, I_1 :

$$
I^{UB} = I^1 := \Lambda(I_1) = [1,1,1,1,1,1] = I_1
$$

So that $I^{LB} = I^{UB} = I^{\star}$.

Setup #4: Country Geography and SCD-C From Above We set the exogenous return in equation (OA.10) equal to $A_i = -4\,\forall\, i \in [1,2]$ and $A_i = -2\,\forall\, i \in [3,4]$ and $A_i = 0\,\forall\, i \in [5,6]$. We also assume the return function exhibits SCD-C from above with $\gamma = -1$. But then applying the mapping from equation (OA.9) three times to the vector of all zeros, I_0, identifies the least element of the set of fixed points of the mapping:

$$
\begin{aligned}
I^1 &:= \Lambda(I_0) = [0,0,0,0,1,1] \\
I^2 &:= \Lambda(I^1) = [0,0,0,0,0,0] \\
I^3 &:= \Lambda(I^2) = [0,0,0,0,1,1] \\
I^{LB} = I^4 &:= \Lambda(I^3) = [0,0,0,0,0,0] = I^2
\end{aligned}
$$

so that a fixed edge is found after three iterations. Starting from the vector of all 1s, I_1 :

$$
\begin{aligned}
I^1 &:= \Lambda(I_1) = [1,1,1,1,1,1] \\
I^2 &:= \Lambda(I^1) = [0,0,0,0,1,1] \\
I^3 &:= \Lambda(I^2) = [0,0,0,0,0,0] \\
I^{UB} = I^4 &:= \Lambda(I^3) = [0,0,0,0,1,1] = I^2
\end{aligned}
$$

So that again we find a fixed edge. In fact, the upper and lower bound are part of the same fixed edge, so that $I^{LB} \neq I^{UB}$ and $I^{LB} \leq I^{\star} \leq I^{UB}$. The profit function maximizing vector must be one of the two vectors enveloped by the fixed edge: $[0,0,0,0,1,0]$ or $[0,0,0,0,0,1]$. We know that both of them have to be fixed points, not fixed edges.

Discussion Three important performance metrics for the squeezing step are the following: (1) how often it needs to be applied before convergence, (2) how tight the resulting bounds are, and (3) whether the bounds are fixed edges or fixed points. Intuitively, tighter bounds mean less additional work in identifying the optimal vector. Likewise, if the bounds are fixed edges, there is no chance for the bounds to coincide with the optimal vector. We can evaluate the above example along these three criteria.

The comparison within geography types yields the following insights: (1) With SCD-C from below, the mapping converges faster. Intuitively, with SCD-C from below, all changes in entries are in "one direction," whereas for SCD-C from above, entries change back and forth. (2) With SCD–C from below, the mapping converges to tighter bounds. (3) With SCD-C from below, fixed edges never occur. Fixed edges are a nuisance since they are not fixed points but can prevent the mapping from reaching a fixed point if it gets stuck in a fixed edge. Furthermore, realizing a point is a fixed edge takes two iterations.

The comparison within complementarity types yields the following insights: (1) City-type geographies converge faster for both types of SCD-C. Intuitively, the clear location hierarchy makes it easier to determine which locations are profitable. (2) City-type geographies have tighter bounds. (3) Country-type geographies are more likely to lead to fixed edges since they have several equally valued locations, not all of which are included in the optimal vector, leading the iteration to result in a fixed edge.

OA.8 Model Fit and Counterfactual Results for Alternative Model Calibrations

In this section, we present results for alternative calibrations of the model. We show the calibration outcomes and counterfactuals for two version of the models calibrated with $\theta = 7.5$ and $\theta = 2.5$. We also present the calibration and counterfactual analysis for a version of our model without fixed costs.

OA.8.1 Alternative Values for θ

In this section, we present the model fit of two alternative calibrations of the model which use different values for the dispersion ($1/\theta$) of location-input-specific productivity shocks. We also present the main results of the counterfactual exercises using these alternative calibrations of our model.

Calibration In addition to our baseline specification, we re-estimate the model with alternative values of θ, as discussed in Section 5.2. Our first alternative value of θ comes from Head and Mayer (2019), who set $\theta = 7.5$ which leads to an even stronger negative complementarity than in our baseline calibration. The second alternative value for θ we use is $\theta = 2.5$ which we include to show a calibration that implies positive complementarities among production locations. In both calibrations, we match all targeted moments exactly. Table OA.3 reports the estimated model's fit of a set of untargeted moments.Figures (OA.5), (OA.6), and (OA.7) show additional measures of fit of the model.

Overall, the calibration with $\theta = 7.5$ exhibits a similar fit as the baseline calibration. The fit of the inward affiliate shares is somewhat better than in our baseline calibration, but the baseline calibration matches the outward trade (export) shares and outward MP (foreign affiliate production) shares better. The calibration with $\theta = 2.5$ performs worse than the baseline calibration on the inward affiliate shares.

Counterfactuals In this section, we replicate all the main graphs from the counterfactual section using the two alternative calibrations of the model that draw on different values of θ. Figure OA.8 shows the real wage responses in our Brexit counterfactual.The top panel show the results for the model calibrated with $\theta = 7.5$ and the bottom panel for the calibration where $\theta = 2.5$. Figure OA.9 shows the total and percentage changes in the count of affiliates in our Brexit counterfactual.The top two panels show the results for the model calibrated with $\theta = 7.5$ and the bottom two panels for the calibration where $\theta = 2.5$. Figure OA.10 shows the real wages responses to the sanctions war.The top panel show the results for the model calibrated with $\theta = 7.5$ and the bottom panel for the calibration where $\theta = 2.5$. Figure OA.11 shows the affiliate relocation responses for the sanctions war counterfactual. The top two panels show the results for the model calibrated with $\theta = 7.5$ and the bottom two panels for the calibration where $\theta = 2.5$.

OA.8.2 The Model without Fixed Costs

In this section, we present the model fit of the calibration of our model without fixed costs. We also present the main results of the counterfactual exercises in the calibrated model without fixed costs.

Calibration Table OA.4 shows the untargeted moments. Figure OA.12 shows the trade and inward MP shares in the data and model. Figure OA.13 shows the same histogram for trade costs and MP costs as in the main body of the paper; since fixed costs are set to zero they do not appear in the graph. Figure OA.14 shows the calibrated technology and fixed cost parameters for the model without fixed costs.

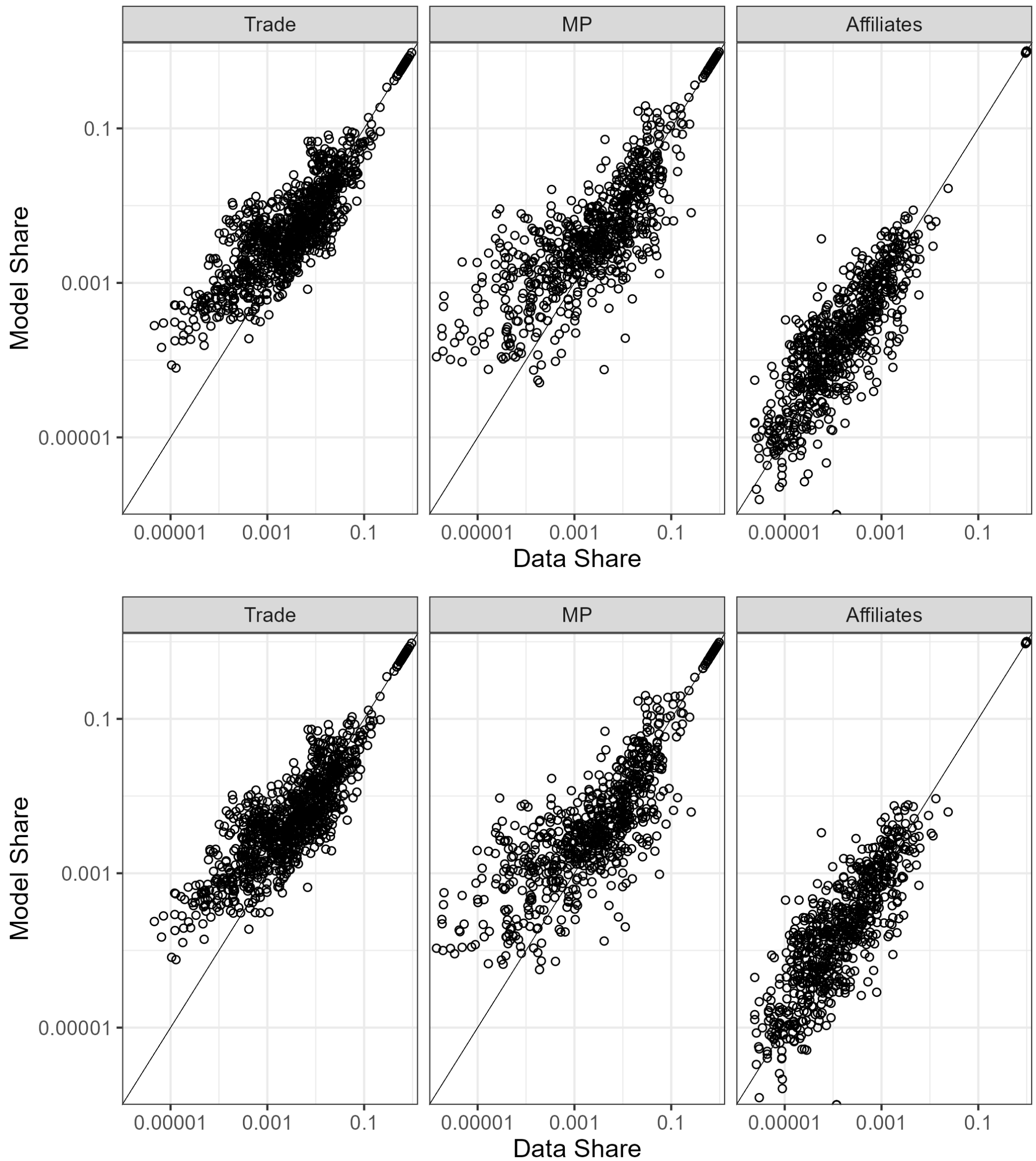
Counterfactuals Figure OA.15 shows the impact of Brexit on real wages in all countries in the model calibration without fixed costs. Figure OA.16 shows the total and percentage changes in the count of affiliates in the Brexit counterfactual in the version of the model without fixed costs. Figure OA.17 shows the change in the number of affiliates, in thousands, in the counterfactual in which we impose sanctions on Russia in the model calibration without fixed costs. Figure OA.18 shows the impact of sanctions on Russia in the model calibration without fixed costs.

Table OA.3: Untargeted Moments for Alternative θ Calibrations

	$\theta = 7.5$				$\theta = 2.5$			
	Inward		Outward		Inward		Outward	
	Data	Model	Data	Model	Data	Model	Data	Model
Panel A: Foreign Trade Shares								
Mean	0.011	0.011	0.010	0.011	0.011	0.011	0.010	0.011
SD	0.023	0.020	0.023	0.021	0.023	0.020	0.023	0.021
Corr	0.822		0.817		0.823		0.824	
R^2	0.676		0.667		0.677		0.678	
Panel B: Foreign MP Shares								
Mean	0.009	0.009	0.006	0.006	0.009	0.009	0.006	0.006
SD	0.026	0.027	0.015	0.015	0.026	0.027	0.015	0.016
Corr	0.728		0.714		0.721		0.726	
R^2	0.530		0.510		0.519		0.528	
Panel C: Foreign Affiliate Shares								
Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SD	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Corr	0.792		0.830		0.682		0.835	
R^2	0.628		0.690		0.465		0.697	

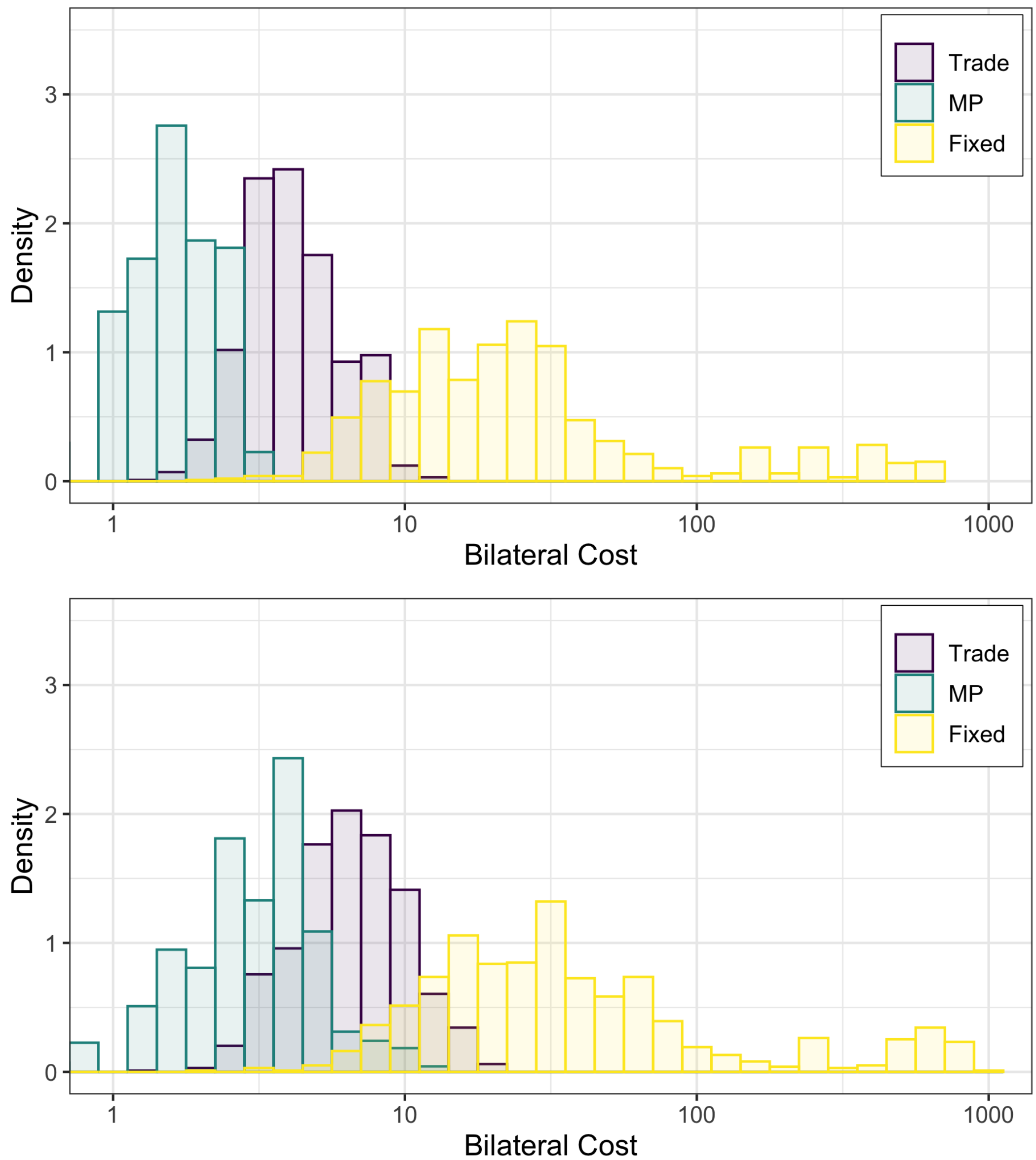
Notes: The figure shows various moments not targeted in the calibration in both calibrated model and data for two different calibrations one using $\theta = 7.5$ the other $\theta = 2.5$. To compute all moments, we drop the diagonals of all bilateral matrices. While we target the coefficients on gravity variables Table 1, we do not directly target the full matrix of foreign trade, MP, or affiliate shares. The inward shares are such that they sum to 1 if added by destination, across origin. For example, inwards trade shares break down each destination market's imports by import origin. Outwards shares sum to 1 if added by origin, across destination. For example, outwards trade shares break down each production country's exports by export destination.

Figure OA.5: Trade Shares, Inward MP Shares, and Inward Affiliate Shares in Data and Model for Alternative θ Calibrations (Top: $\theta = 7.5$; Bottom: $\theta = 2.5$)



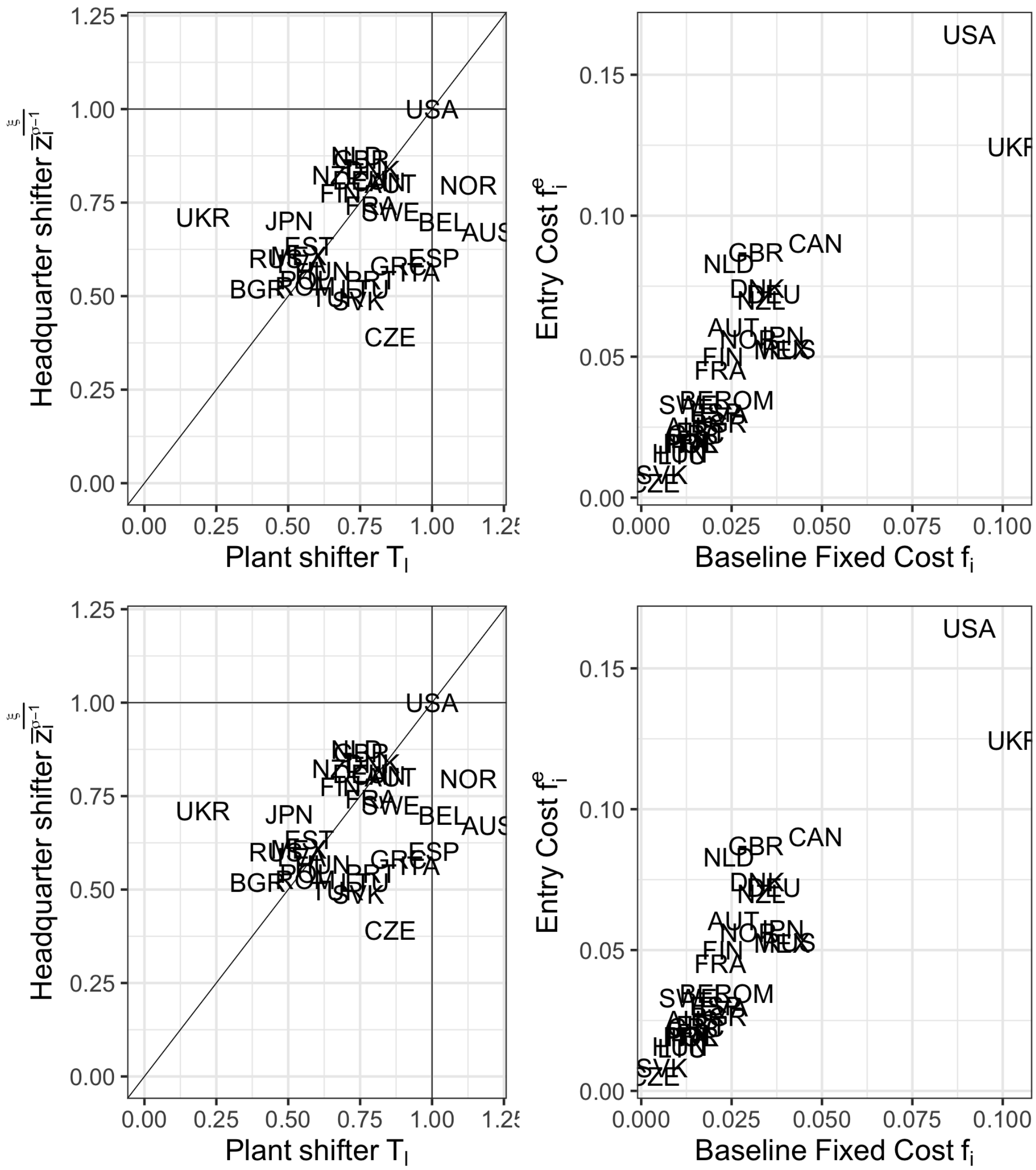
Notes: The figures shows graphs statistics from the data obtained from Alviarez (2019) against the same objects in two alternative calibrations of the model, one for $\theta = 7.5$ the other for $\theta = 2.5$. The top panel shows output from a model calibrated using $\theta = 7.5$, the bottom panel shows output from a calibration with $\theta = 2.5$. In each row, the left panel shows trade shares, the second panel shows inward MP shares, and the third panel shows inward affiliate shares.

Figure OA.6: Trade Costs, MP Costs, and the Bilateral Component Fixed Cost for Alternative θ Calibrations(Top: $\theta = 7.5$; Bottom: $\theta = 2.5$)



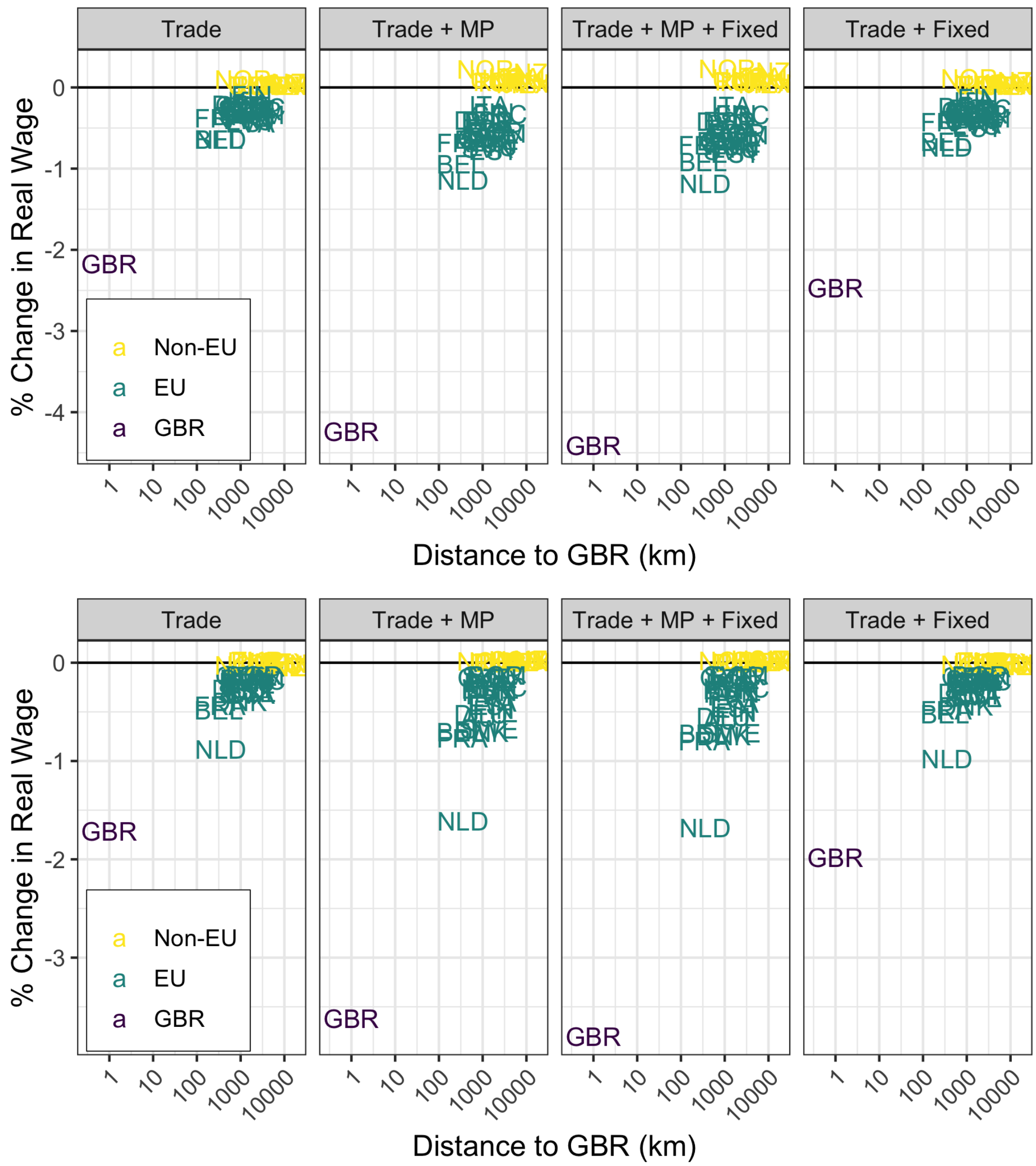
Notes: The figure shows a histogram of our calibrated MP costs, trade costs, and the bilateral component of fixed costs in two alternative calibrations of the model. The top panel shows values for a calibration that uses $\theta = 7.5$, the bottom panel for a calibration that uses $\theta = 2.5$. We omit the own-costs which are normalized to 1 for all three types of costs. For MP and fixed costs, we also omit country pairs where MP is zero, since we set the MP costs to be infinity in those cases.

Figure OA.7: Technology, the Base Component of Fixed Costs, and Entry Costs for Alternative θ Calibrations (Top: $\theta = 7.5$; Bottom: $\theta = 2.5$)



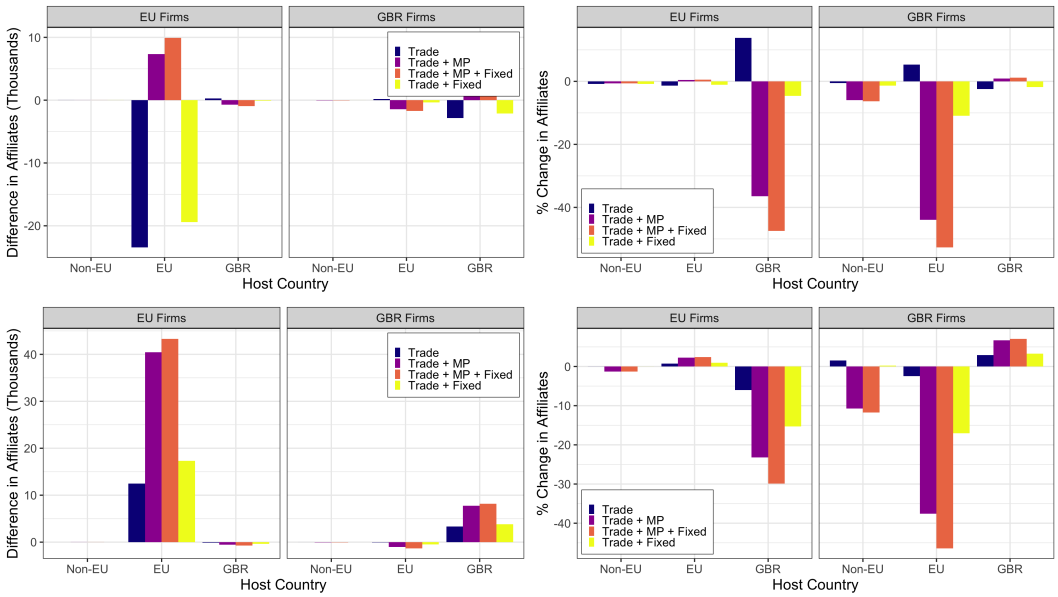
Notes: The figure shows a number of calibrated shifters in alternative calibrations of the model, one with $\theta = 7.5$ and one with $\theta = 2.5$. The top panels are form a calibration using $\theta = 7.5$, the bottom panel from a calibration using $\theta = 2.5$. The left panel in each row graphs the Pareto minimum $\underline{z}_i^{\xi/(\sigma-1)}$ of the firm productivity distribution against the scale T_ℓ of the Fréchet distribution of location-input-specific productivity shocks. The terms $\underline{z}_i^{\xi/(\sigma-1)}$ and T_ℓ appear multiplicatively in the expression for trilateral flows. The right panel in each row plots the entry cost f_i^e against the base component of the fixed cost f_i.

Figure OA.8: Real Wages Changes Across Countries in the Brexit Simulation for Alternative θ Calibrations (Top: $\theta = 7.5$; Bottom: $\theta = 2.5$)



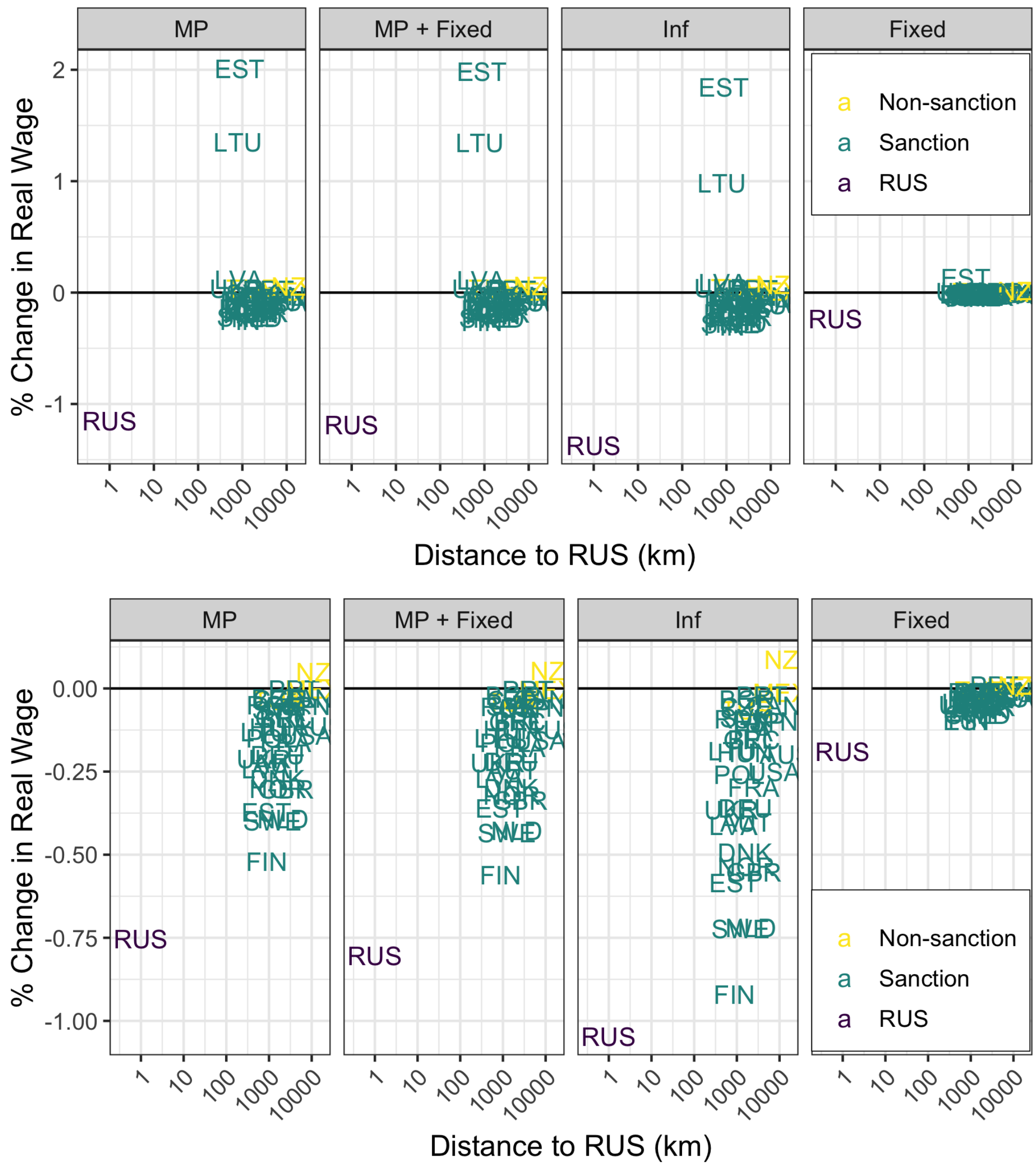
Notes: The figures shows the change in real wage in our Brexit simulation for two alternative calibrations one with $\theta = 7.5$ and one with $\theta = 2.5$. The top panel is for the case of $\theta = 7.5$, the bottom panel for the case of $\theta = 2.5$. The effect on EU member countries is shown in green, the effect on non-EU countries is shown in yellow, and the effect on Great Britain is shown in purple. In the first panel, only trade costs increase by 10%, the second panel adds a 10% increase of the MP costs, and the third adds a 10% increase in the fixed costs. The fourth panel considers the trade and fixed cost increases in isolation.

Figure OA.9: Net (Percentage) Changes in Affiliate Counts by Sender and Host Country in the Brexit Simulation for Alternative θ Calibrations (Top: $\theta = 7.5$; Bottom: $\theta = 2.5$)



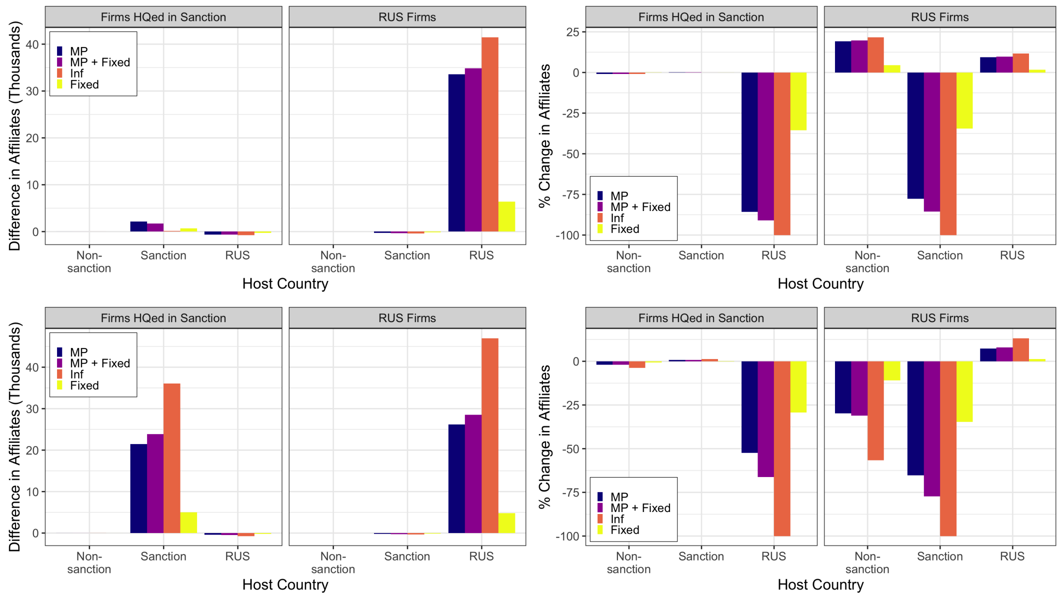
Notes: The left panels of the figure shows the absolute change and percentage change (top and bottom) in the number of affiliates operated by firms headquartered in the EU in non-EU countries, EU countries, and in Great Britain in the Brexit simulation for two alternative calibrations of the model one with $\theta = 7.5$ and one with $\theta = 2.5$. The right panel shows the change in the number of affiliates operated by firms headquartered in Great Britain in non-EU countries, EU countries, and in Great Britain.The top panels are for the case of $\theta = 7.5$, the bottom panels are for the case of $\theta = 2.5$. The blue bar reflects a 10% trade costs increase, the purple bar adds a 10% increase of the MP costs, the orange bar adds a 10% increase in fixed costs. The yellow bar considers the trade and fixed cost increases in isolation.

Figure OA.10: Real Wages Changes Across Countries in the Sanctions on Russia Simulation for Alternative θ Calibrations (Top: $\theta = 7.5$; Bottom: $\theta = 2.5$)



Notes: The figures shows the change in real wage in our Sanctions on Russia simulation for two alternative calibrations one with $\theta = 7.5$ and one with $\theta = 2.5$. The top panel is for the case of $\theta = 7.5$, the bottom panel for the case of $\theta = 2.5$. The effect on countries imposing sanctions on Russia is shown in green, the effect on non-sanctioning countries is shown in yellow, and the effect on Russia is shown in purple. In the first panel, only MP costs increase by 30% while the second panel adds a 30% increase of the fixed costs. In the third panel, we set MP cost to infinity. The fourth panel shows the case in which only fixed costs increase by 30%relative to the baseline.

Figure OA.11: Net (Percentage) Changes in Affiliate Counts by Sender and Host Country in the Sanctions on Russia Simulation for Alternative θ Calibrations (Top: $\theta = 7.5$; Bottom: $\theta = 2.5$)



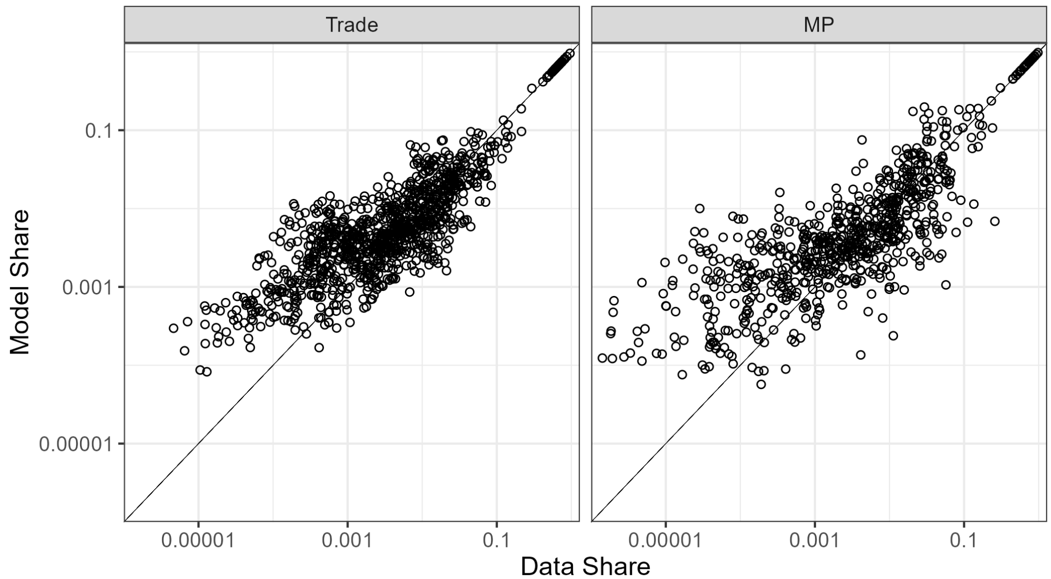
Notes: The figure shows the total change and percentage change in the number of affiliates operated by firms headquartered in countries that placed sanctions to Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries for two alternative calibrations of the model one with $\theta = 7.5$ and one with $\theta = 2.5$. The top panels are for the case of $\theta = 7.5$, the bottom panels are for the case of $\theta = 2.5$. The leftmost panel in each row of the figure shows the change in the number of affiliates operated by firms headquartered in countries that placed sanctions to Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries. The second panel from the left in each row of the figure shows the change in the number of affiliates operated by firms headquartered in Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries. The third panel and fourth panel in each row of the figure show the same changes in percentage terms. The blue bar refers to a counterfactual with a 30% increase in the cost of MP, the purple bar adds a 30% increase in the fixed costs, the fourth (yellow) bar sets the MP cost to infinity.

Table OA.4: Untargeted Moments for the Calibration Without Fixed Costs

	Inward		Outward	
	Data	Model	Data	Model
Panel A: Foreign Trade Shares				
Mean	0.011	0.011	0.010	0.011
SD	0.023	0.019	0.023	0.021
Corr	0.833		0.823	
R^2	0.693		0.677	
Panel B: Foreign MP Shares				
Mean	0.009	0.009	0.006	0.006
SD	0.026	0.027	0.015	0.015
Corr	0.721		0.727	
R^2	0.520		0.528	
Panel C: Foreign Affiliate Shares				
Mean	0.000	0.031	0.000	0.031
SD	0.001	0.084	0.001	0.025
Corr	0.284		0.094	
R^2	0.081		0.009	

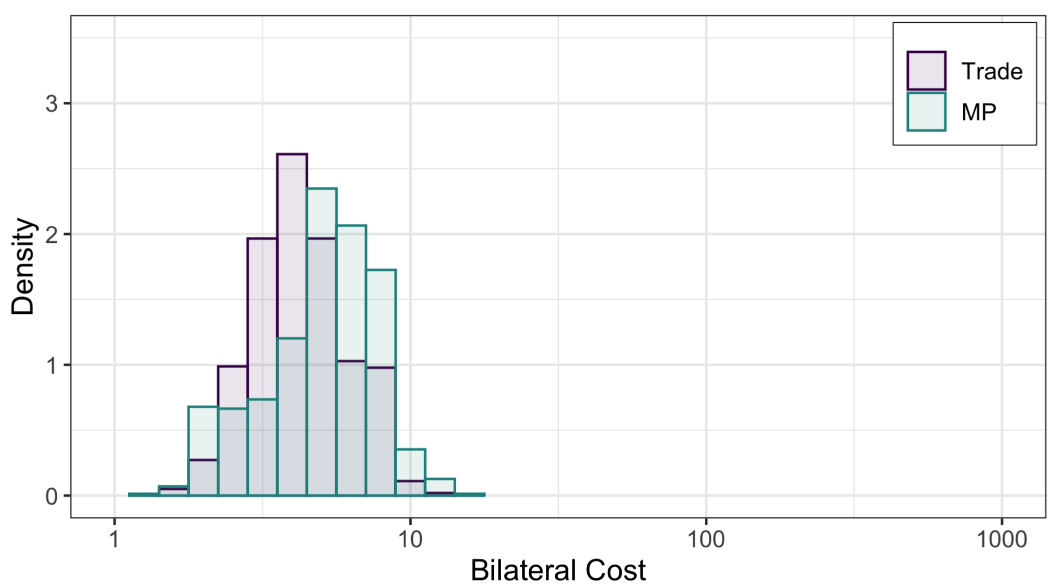
Notes: The figure shows various moments not targeted in the calibration in both calibrated model and data for the calibration of the model without fixed costs. To compute all moments, we drop the diagonals of all bilateral matrices. While we target the coefficients on gravity variables Table 1, we do not directly target the full matrix of foreign trade, MP, or affiliate shares. The inward shares are such that they sum to 1 if added by destination, across origin. For example, inwards trade shares break down each destination market's imports by import origin. Outwards shares sum to 1 if added by origin, across destination. For example, outwards trade shares break down each production country's exports by export destination.

Figure OA.12: Trade Shares and Inward MP Shares in Data and Model for the Calibration Without Fixed Costs



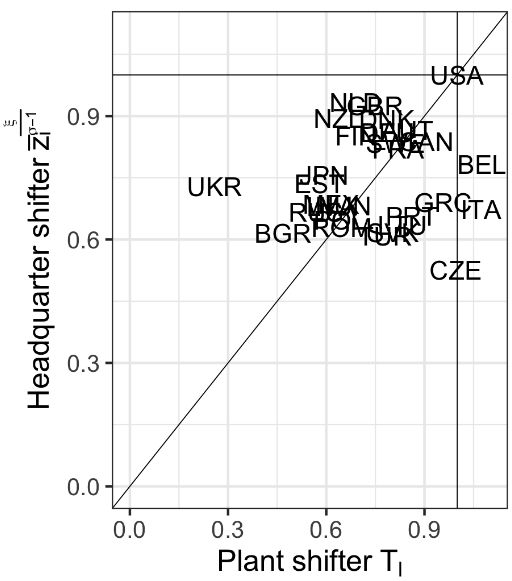
Notes: The figures shows graphs statistics from the data obtained from Alviarez (2019) against the same objects from the calibration of the model without fixed costs. The left panel shows trade shares, the second panel shows inward MP shares, and the third panel shows inward affiliate shares.

Figure OA.13: Trade Costs and MP Costs for the Calibration Without Fixed Costs



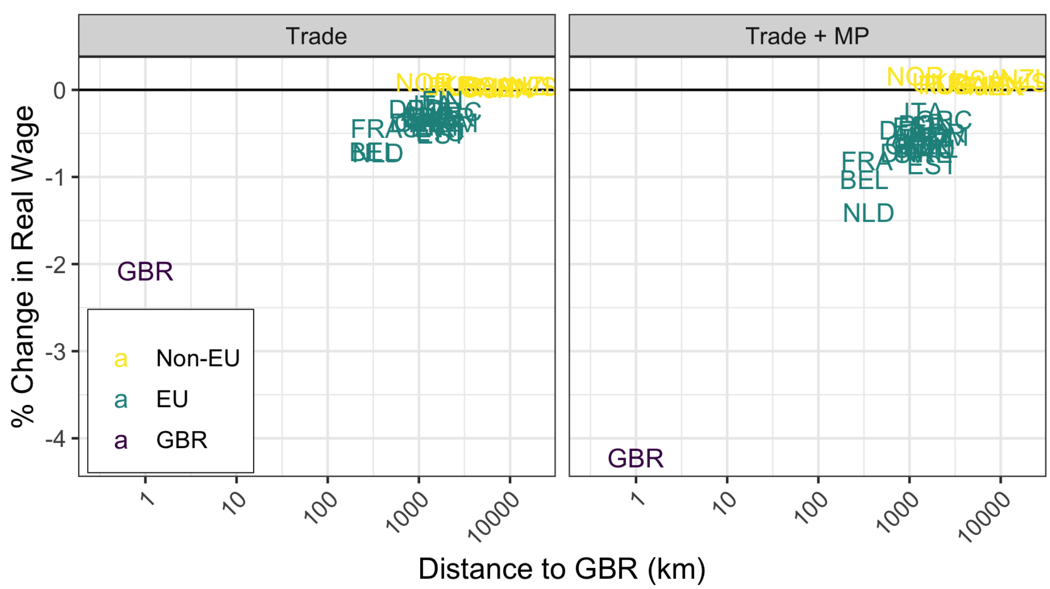
Notes: The figure shows a histogram of the calibrated MP costs and trade costs in the model without fixed costs. We omit the own-costs which are normalized to 1 for all three types of costs. For MP costs, we also omit country pairs where MP is zero, since we set the MP costs to infinity in those cases.

Figure OA.14: Technology Parameters for the Calibration Without Fixed Costs



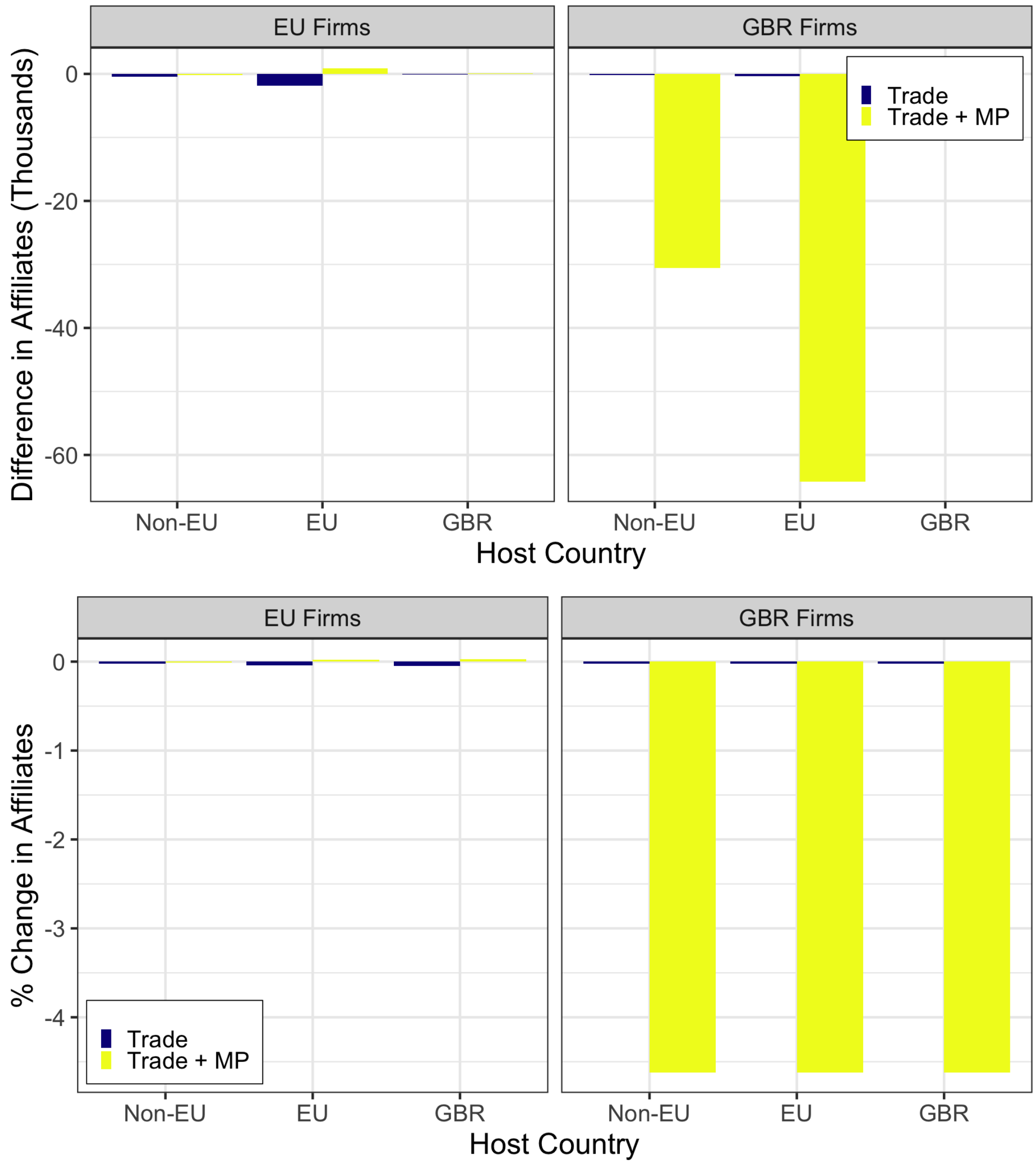
Notes: The estimated country-specific fundamentals of the calibrated model without fixed costs. The left panel plots the Pareto minimum $\underline{z}_i^{\frac{\xi}{\sigma-1}}$ of the firm productivity distribution against the scale T_ℓ of Fréchet distribution of location-input-specific productivity shocks.

Figure OA.15: Real Wages Changes Across Countries in the Brexit Simulationfor the Calibration Without Fixed Costs



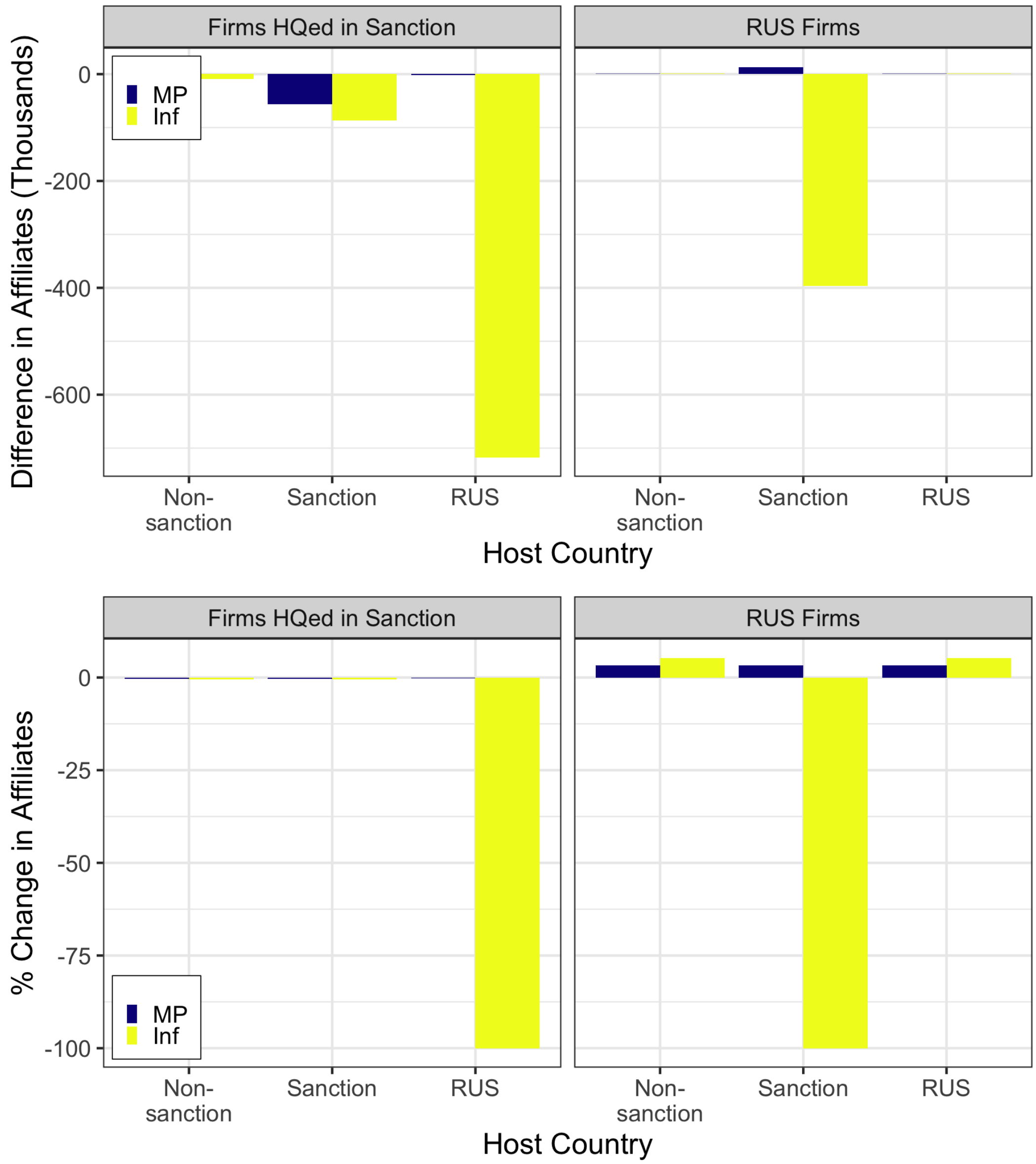
Notes: The figures shows the change in real wage in our Brexit simulation for the calibrated model without fixed costs. The effect on EU member countries is shown in green, the effect on non-EU countries is shown in yellow, and the effect on Great Britain is shown in purple. In the first panel, only trade costs increase by 10% and the second panel adds a 10% increase of the MP costs.

Figure OA.16: Net (Percentage) Changes in Affiliate Counts by Sender and Host Country in the Brexit Simulation for the Calibration Without Fixed Costs



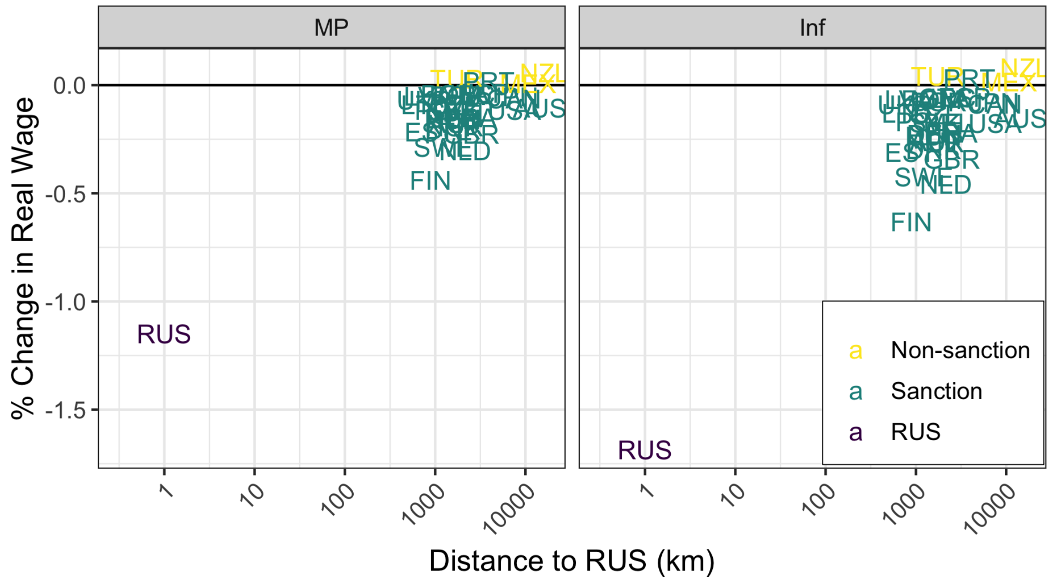
Notes: The left panel in each row of the figure shows the absolute and percentage change (top and bottom) in the number of affiliates operated by firms headquartered in the EU in non-EU countries, EU countries, and in Great Britain in the Brexit simulation in the calibrated model without fixed costs. The right panel in each row of the figure also shows the change in the number of affiliates operated by firms headquartered in Great Britain in non-EU countries, EU countries, and in Great Britain. The blue bar reflects a 10% trade costs increase, the yellow bar adds a 10% increase of the MP costs.

Figure OA.17: Net (Percentage) Changes in Affiliate Counts by Sender and Host Country in the Sanctions on Russia Simulation for the Calibration Without Fixed Costs



Notes: The left panel in each row of the figure shows the absolute and percentage changes (top and bottom) in the number of affiliates operated by firms headquartered in countries that placed sanctions to Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries, for the calibration of our model without fixed costs. The right panel in each row of the figure shows the change in the number of affiliates operated by firms headquartered in Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries. The blue bar refers to a counterfactual with a 30% increase in the MP cost while the yellow bar reflects the effect of setting the MP cost to infinity.

Figure OA.18: Real Wages Changes Across Countries in the Sanctions on Russia Simulation for the Calibration Without Fixed Costs



Notes: The figures shows the change in real wage in our Sanctions on Russia simulation for the calibrated model without fixed costs. The effect on EU member countries is shown in green, the effect on non-EU countries is shown in yellow, and the effect on Great Britain is shown in purple. In the first panel, only trade costs increase by 10% and the second panel adds a 10% increase of the MP costs.