NBER WORKING PAPER SERIES

COMBINATORIAL DISCRETE CHOICE:
THEORY AND APPLICATION TO MULTINATIONAL PRODUCTION

Costas Arkolakis
Fabian Eckert
Rowan Shi

Working Paper 31877
http://www.nber.org/papers/w31877

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2023, Revised May 2025

We are indebted to invaluable help with the Julia code from Junyuan Chen. We are grateful to Pablo Fajgelbaum and Eduardo Morales for excellent discussions at the Summer Trade Workshop at Princeton IES. We thank Jake DiCicco, Tra Nguyen, Andrew Salmon, Benjamin Tong, Yunus Tuncbilek, and Nardeen Abdulkareem for outstanding research assistance. We also thank Chrysanthos Gounaris, David Atkin, Arnaud Costinot, Dave Donaldson, Penny Goldberg, Navin Kartik, Samuel Kortum, Nitya Pandalai-Nayar, and Steve Redding for comments and suggestions. Shi gratefully acknowledges support from the Princeton IES. Code implementing the techniques introduced in this paper can be found at https://github.com/rowanxshi/CDCP.jl. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Combinatorial Discrete Choice: Theory and Application to Multinational Production
Costas Arkolakis, Fabian Eckert, and Rowan Shi
NBER Working Paper No. 31877
November 2023, Revised May 2025
JEL No. F12, F21, F6

ABSTRACT

Discrete choice problems with complementarities among options quickly grow infeasible to solve, since they generically require evaluating all combinations of choices. We develop a solution method that applies whenever choices are weakly complementary or substitutable, using the implied choice monotonicity to discard suboptimal combinations without computing their payoff. It is orders of magnitude faster than existing approaches, finds the global solution, and extends to heterogeneous-agent settings. Using our method, we calibrate a general equilibrium model of multinational firms selecting global production locations to show that complementarities among locations can amplify, dampen, or even reverse the welfare gains from multinational production.

Costas Arkolakis
Yale University
Department of Economics
and NBER
costas.arkolakis@yale.edu

Fabian Eckert
University of California, San Diego
f@fpeckert.me

Rowan Shi
Toronto Metropolitan University
rowanxshi@torontomu.ca

1. Introduction

Complementarities among discrete choices typically force decision-makers to evaluate all possible *combinations*. The resulting dimensionality renders such combinatorial discrete choice problems practically infeasible to solve without additional structure—even with a moderate number of options. Yet, complementarities arise naturally from network, scale, or cannibalization effects in economic settings, such as multinational firms choosing where to operate or retailers selecting store locations.

We develop a unified method to solve combinatorial discrete choice problems (CDCPs) whose objective functions satisfy a simple single-crossing condition, enforcing a weak form of complementarity or substitutability. Our approach extends to heterogeneous-agent settings where payoffs depend on type, is many orders of magnitude faster than existing methods, and always finds the global optimum. We apply it to a quantitative general equilibrium model of multinational production, in which heterogeneous firms choose global production locations. The firm problem is a CDCP: complementarities arise from scale and cannibalization effects, and fixed costs make location choices discrete. Using the calibrated model, we show that incorporating complementarities and fixed costs—enabled by our method—materially affects the estimated welfare gains from multinational production.

Our method builds on Jia (2008), which develops the insight that positive complementarities induce a form of *choice monotonicity* that can be used to iteratively discard suboptimal choice combinations without computing their payoff. If an option yields a negative payoff even when all other options are selected—maximizing complementarities—it cannot be part of the optimal combination. Likewise, if an option has a positive payoff even in isolation, it must be included.

We show that the elimination logic in Jia (2008) extends beyond the case of positive complementarities to problems satisfying a weaker single-crossing condition. This condition requires only that if a choice yields a positive payoff as part of a given combination, its payoff remains weakly positive either when additional choices are added or when existing ones are removed. Notably, our single crossing condition nests the case of negative complementarities—for example, when new locations cannibalize demand—for which few solution methods exist, and to which the elimination logic was previously assumed not to apply.

Iteratively applying the elimination logic often dramatically reduces the set of potentially optimal combinations, but may still leave multiple candidates. In such cases, Jia (2008) applies the brute force approach of explicitly computing the payoff of all remaining combinations. Instead, we introduce a recursive branching procedure that reapplies the elimination logic to subsets of the remaining candidates until the optimal combination is found. Our branching method collapses to brute force in the worst case only.

CDCPs with heterogeneous agents pose an added challenge: payoffs vary by agent type, so the solution for one type need not apply to others. The incumbent approach—discretizing the type space and solving a CDCP at each grid point—may introduce interpolation error. We further extend our method to recover the exact *policy function* mapping types to optimal combinations of choices. The extension requires a second single-crossing condition that induces a form of *type monotonicity*, allowing us to identify the cutoff types where the policy function changes value without solving the CDCP for every type.

We apply our method to solve a general equilibrium model of multinational production where heterogeneous firms select locations for production. The firm problem is a CDCP: scale economies and demand cannibalization across production locations give rise to complementarities, while location-specific fixed costs make location choices discrete. In addition, firm heterogeneity implies that optimal location combinations differ across firms. We show how our single crossing conditions correspond to explicit parameter restrictions in the model.

We calibrate the model to match trade and multinational activity across 32 countries. Because production location sets vary across firms, the model does not yield the aggregate log-linear estimating equations used in standard gravity-based calibrations. Instead, we solve the full model repeatedly to match aggregate flows, computing the policy function that maps firm productivity to optimal location choices at each iteration. Our method makes this strategy feasible even with many countries, arbitrary firm heterogeneity, and a wide range of complementarities—from strongly positive to strongly negative.

We benchmark the performance of our method using up to 256 synthetic countries generated by sampling from the empirical distributions of productivity and cost parameters obtained from our calibration. With positive complementarities, our method is up to four orders of magnitude faster than brute-force and an order of magnitude faster than the method in Jia (2008) with discretized firm heterogeneity. It performs equally fast with negative complementarities for which brute force was the only incumbent solution method. The speed of our method matters especially in general equilibrium settings, where policy functions must be solved repeatedly.

To avoid the computational burden of solving CDCPs, many quantitative models of multinational production rule them out by abstracting from mechanisms that give rise either to complementarities, such as scale economies or cannibalization, or to discrete choices, such as fixed costs. We assess the role of these ingredients in shaping the counterfactual welfare predictions of the full model. To guide interpretation, we derive a new expression for the gains from multinational production that extends the welfare formula in Arkolakis, Costinot, and Rodríguez-Clare (2012) to our setting with scale economies, cannibalization, and fixed costs. Our formula decomposes the gains from multinational production into intuitive channels, including one that captures the effective productivity gains unlocked by firms that operate multiple global production locations.

Quantitatively, both complementarities and fixed costs shape the gains from multinational production. Countries with many multinationals benefit significantly, as international production reduces marginal costs and generates large profits. In contrast, less productive countries with few multinationals may lose: few of their firms generate profit abroad, while they play host to the local affiliates of foreign firms that crowd out domestic producers. Positive complementarities amplify this dynamic by creating increasing returns for multinationals, further boosting home-country gains and intensifying the crowding out in host countries. On the other hand, negative complementarities temper this effect. Calibrations that omit fixed costs overstate welfare gains, because all firms participate in the productivity gains accessible through multinational production.

Modeling complementarities and agent heterogeneity has long posed a challenge in the discrete choice literature. The classic random utility framework of McFadden (1974) assumes perfect substitutability, so that each agent chooses a single option. This assumption permits rich multidimensional heterogeneity via option-specific shocks and yields closed-form aggregation. However, the literature has struggled to allow for imperfect substitutability or complementarity among options. Some authors allow agents to choose multiple options at once but without interdependencies in payoffs (see, e.g., Hendel 1999). Others associate each combination of options with a random utility component (see, e.g., Train, McFadden, and Ben-Akiva 1987; Gentzkow 2007), but aggregation then requires evaluating payoffs for all combinations which quickly becomes infeasible in practice.

Our paper contributes to a growing quantitative literature on combinatorial discrete choice problems that abstract from random utility components, allowing instead for large numbers of options, more flexible functional forms, and complementarities. Jia (2008) introduces a foundational method for solving CDCPs with strongly positive complementarities, often applied to store expansion and firm sourcing problems (see Antras, Fort, and Tintelnot 2017; Alfaro-Urena et al. 2023; Antràs et al. 2024b). Our extension to negative complementarities has similarly been applied in recent work (see, e.g., Jiang 2023; Liu 2023; Sabal 2025). By developing a method that handles both positive and negative complementarities, we also lay the groundwork to solve mixed-complementarity problems that arise, for example, in the multinational production model of Antràs et al. (2024b). Castro-Vincenzi et al. (2024) build on our method to solve such problems. Moreover, our heterogeneous-agent method allows the literature to avoid the approximation errors associated with discretizing heterogeneity when solving CDCPs (see Tintelnot 2017; Antràs et al. 2024b).

Our method allows models of multinational production to incorporate cross-location interdependencies and fixed costs without sacrificing tractability. Existing work avoids CDCPs by abstracting from these features (see Ramondo and Rodríguez-Clare 2013; Ramondo 2014; Arkolakis, Ramondo, et al. 2018); solves small CDCPs via brute force (see Zheng 2016; Tintelnot 2017; Dyrda, Hong, and Steinberg 2024); or restricts attention to positive complementarities to apply the method of Jia (2008) (see Antras, Fort, and Tintelnot 2017; Alfaro-Urena et al. 2023; Antràs et al. 2024b). We also show that complementarities and fixed costs are essential for quantifying the gains from multinational production, both in our calibrated model and by extending the sufficient-statistics framework of Arkolakis, Costinot, and Rodríguez-Clare (2012) to environments with these ingredients.

Finally, we also contribute to a growing literature on algorithmic solutions to CDCPs. Recent work imposes linear objective functions to enable integer programming (e.g., Head, Mayer, et al. 2024), adopts greedy algorithms that do not guarantee global optimality (e.g., Fan and Yang 2020), or uses deep learning to approximate policy functions in heterogeneous-agent CDCPs (e.g., Kulesza 2024). In contrast, our method applies to all objectives that satisfy our single crossing conditions, identifies the global optimum, and allows for exact aggregation across heterogeneous agents.

2. A Unified Framework to Solve CDCPs

In this section, we formally define combinatorial discrete choice problems and show how to solve them in cases where decisions satisfy a weak form of complementarity or substitutability. Throughout, L denotes a finite set of items indexed by ℓ; $\mathcal{P}(L)$ its power set, that is, the collection of all its possible subsets; and $\mathcal{L}$ an arbitrary element of the power set. We consider an objective function $f : \mathcal{P}(L) \to \mathbb{R}$ that maps elements from the power set to a real number and denote the space of such functions by $\mathcal{F} = \{f : \mathcal{P}(L) \to \mathbb{R}\}$. The Online Appendix contains a formal mathematical treatment of all statements and results.

2.1. Defining Combinatorial Discrete Choice Problems

Consider a decision-maker choosing a set $\mathcal{L} \in \mathcal{P}(L)$ to maximize an objective function f. Such discrete choice problems frequently appear in economic models, for example when a firm selects countries from which to import or a government chooses where to implement infrastructure projects. Our example throughout this section is a profit-maximizing firm selecting a set of foreign production locations $\mathcal{L}$.

Problems with multiple discrete choices are straightforward to solve when choices are independent, that is, when the valuation of any location $\ell \in L$ does not depend on the composition of the chosen set $\mathcal{L}$. In many settings, however, locations are interdependent: the value of a location ℓ depends on the other locations in $\mathcal{L}$. We introduce the following marginal value operator to formalize the notion of interdependencies among locations:

Definition 1 (Marginal Value Operator)**.** *For any $\ell \in L$ and $\mathcal{L} \in \mathcal{P}(L)$, define the marginal value operator $D_\ell : \mathcal{F} \to \mathcal{F}$ by*

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

If $D_\ell f(\mathcal{L})$ does not vary in $\mathcal{L}$ for each $\ell \in L$, the problem decomposes into $|L|$ independent decisions. If $D_\ell f(\mathcal{L})$ varies with $\mathcal{L}$, locations are interdependent and the firm must consider *combinations* of locations. We now formally define the class of combinatorial discrete choice problems.

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

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

where there is at least one $\ell \in L$ and two sets $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ for which $D_\ell f(\mathcal{L}) \neq D_\ell f(\mathcal{L}')$.

We refer to these types of maximization problems as combinatorial because, generically, solving them requires evaluating the objective function for every element in the set $\mathcal{P}(L)$, which grows exponentially with the number of locations L.

2.2. Solving Combinatorial Discrete Choice Problems

Without additional structure on the objective function, no known polynomial-time algorithm exists for solving CDCPs, making the generic problem NP-hard. However, when the objective function f satisfies a single-crossing property, we show that an iterative solution method applies. In this section, we first introduce the property and then discuss its implications for solving CDCPs.

Single Crossing Differences in Choices The single crossing differences in choices property restricts how the marginal value of a location ℓ varies with the decision set $\mathcal{L}$.

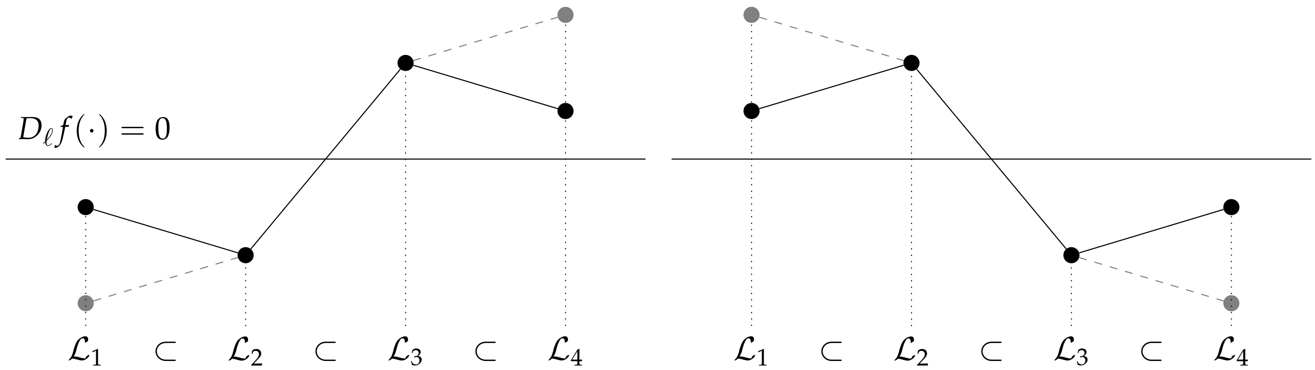
Definition 3 (SCD-C)**.** *The objective function f satisfies single crossing differences in choices from above if, for every $\ell \in L$, and for all decision sets $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ such that $\mathcal{L} \subset \mathcal{L}'$,*

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

and single crossing differences in choices from below if

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

The SCD-C condition restricts the marginal value function of location ℓ to change sign at most once. Thus, the conditions captures a weak form of complementarity, where a location



(A) SCD-C from below

(B) SCD-C from above

FIGURE 1: EXAMPLE MARGINAL VALUE FUNCTIONS

Both panels show the marginal value of location as a function of a succession of nested decision sets. The solid black line depicts a function that satisfies the weaker single crossing condition, but not the stricter super- or submodularity condition. The dashed line depicts a function that satisfies this stronger requirement. Note that under SCD-C, marginal values can increase or decrease arbitrarily as long as they only cross zero once.

with a positive marginal value for some decision set $\mathcal{L}$ retains a positive marginal value as additional locations are added (SCD-C from below) or removed (SCD-C from above) from $\mathcal{L}$. The solid and dashed lines in the left panel of Figure 1 both show examples of how the marginal value of a location ℓ can vary with SCD-C from below as locations are added to an initial location set $\mathcal{L}_1$. While the marginal value may change and even decrease (solid line), it crosses zero at most once from below. The right panel shows two marginal value functions with SCD-C from above, which can cross zero at most once from above.[1]

The well-known properties of supermodularity and submodularity capture a stronger notion of complementarity and serve as sufficient conditions for SCD-C to hold. A function f is submodular if marginal values are monotonically decreasing, that is if $D_\ell f(\mathcal{L}) \geq D_\ell f(\mathcal{L}')$ for all $\mathcal{L} \subseteq \mathcal{L}'$ and any ℓ, which implies SCD-C from above. Supermodularity reverses the inequality and is sufficient for SCD-C from below. While both the solid and dashed lines in Figure 1 show marginal value functions consistent with single crossing differences in choices, only the dashed lines are consistent with sub- or supermodularity.

The weaker form of complementarity implied by the SCD-C condition broadens the range of economic problems to which our methods apply. Consider a multinational firm like Ford, which profitably operates plants in Germany and the United States. Supermodularity requires

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

that adding a plant in Canada raises the marginal value of both the German and U.S. plants. In reality, however, the Canadian plant may lower the value of the U.S. plant—by cannibalizing its sales to the Canadian market—while increasing the value of the German plant through enhanced global scale economies. Although such a scenario violates super- and submodularity, it remains consistent with SCD-C from below, as long as the U.S. plant's marginal value stays weakly positive when the Canadian plant is added.

The Squeezing Procedure: Reducing a CDCP's Domain We now introduce a method for solving combinatorial discrete choice problems when the objective function satisfies either form of the SCD-C condition.

Our method iteratively shrinks the domain of the CDCP without eliminating the optimal decision set $\mathcal{L}^\star \equiv \arg\max_{\mathcal{L}\in\mathcal{P}(L)} f(L)$. To operationalize this approach, we introduce the notion of a "bounding pair," a pair of sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ which defines a restricted subdomain of the original CDCP that always contains $\mathcal{L}^\star$, so that $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}$. The full domain corresponds to the "trivial" bounding pair $[\emptyset, L]$. We refer to $\underline{\mathcal{L}}$ as the lower bounding set and $\overline{\mathcal{L}}$ as the upper bounding set, and say that $[\underline{\mathcal{K}}, \overline{\mathcal{K}}]$ is "tighter" than $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ if $\underline{\mathcal{L}} \subseteq \underline{\mathcal{K}}$ and $\overline{\mathcal{K}} \subseteq \overline{\mathcal{L}}$.

We conceptualize solving a CDCP as eliminating all non-optimal decision sets from the problem's domain. We thus introduce a mapping with the goal of tightening a bounding pair around the optimal decision set $\mathcal{L}^\star$, progressively "squeezing" non-optimal sets out of the domain. We refer to this mapping as the squeezing step.

Definition 4 (Squeezing step). *The squeezing step is the mapping $S : \mathcal{P}(L) \times \mathcal{P}(L) \to \mathcal{P}(L) \times \mathcal{P}(L)$ defined for any $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ as:*

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

where $\Phi : \mathcal{P}(L) \to \mathcal{P}(L)$ is defined as

$$\Phi\left(\mathcal{L}\right) = \left\{\ell \in L \mid D_\ell f\left(\mathcal{L}\right) > 0\right\}.$$

When the objective function satisfies SCD-C, iteratively applying the squeezing step to the trivial bounding pair $[\emptyset, L]$ generates a sequence of progressively tighter bounding pairs, converging in at most $|L|$ steps. We now provide a constructive proof of this claim before formally stating the result in Theorem 1.

The squeezing step builds on the mapping Φ, which identifies the set of locations $\ell \in L$ with positive marginal value given the decision set $\mathcal{L}$. Importantly, $\Phi(\mathcal{L}^\star) = \mathcal{L}^\star$ by construction, highlighting that Φ is the discrete analogue of a first order condition. Section 2.3 provides a detailed discussion of Jia (2008), which first introduced Φ in the context of a supermodular objective function.

The mapping Φ is monotone if and only if the underlying objective function f satisfies the SCD-C condition. The type of SCD-C determines the direction of the monotonicity. Consider two decision sets $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(L)$ such that $\mathcal{L} \subset \mathcal{L}'$. With SCD-C from above, if a location ℓ has a positive marginal value in the larger set $\mathcal{L}$, it must also have a positive marginal value in any nested set $\mathcal{L}'$. As a result, $\mathcal{L} \subset \mathcal{L}'$ implies $\Phi(\mathcal{L}') \subseteq \Phi(\mathcal{L})$: the mapping Φ is order-reversing. With SCD-C from below, the logic is inverted: if ℓ has a positive marginal value as part of the smaller set $\mathcal{L}$, then it must have positive marginal value as part of the larger set. Then, $\mathcal{L} \subset \mathcal{L}'$ implies $\Phi(\mathcal{L}) \subseteq \Phi(\mathcal{L}')$ and the mapping Φ is order-preserving.[2]

The monotonicity of Φ simplifies the mapping S when we know the type of SCD-C the objective function f satisfies. With SCD-C from above, the mapping Φ is order-reversing, so that applying Φ to the lower bounding set must return a decision set that is always nested in the decision set that results from applying Φ to the upper bounding set. As a result, $S\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right) = \left[\Phi\left(\overline{\mathcal{L}}\right), \Phi\left(\underline{\mathcal{L}}\right)\right]$. With SCD-C from below, the reverse logic simplifies the squeezing step to $S\left(\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]\right) = \left[\Phi\left(\underline{\mathcal{L}}\right), \Phi\left(\overline{\mathcal{L}}\right)\right]$.

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

Importantly, the monotonicity of Φ guarantees that, starting from the trivial bounding pair $[\varnothing, L]$, iteratively applying the squeezing step produces bounding pairs that weakly tighten with each iteration. For a concrete example, suppose f satisfies SCD-C from above, so that Φ is order-reversing. Then applying the squeezing step to the trivial bounding pair yields

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

Note that the new bounding $[\Phi(L), \Phi(\varnothing)]$ is vacuously tighter than the trivial bounding pair $[\varnothing, L]$. Applying the squeezing step again produces:

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

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

The order-reversing property of Φ guarantees that the new upper bounding set $\Phi(\Phi(L))$ is (weakly) tighter than the previous upper bounding set $\Phi(\emptyset)$: note $\emptyset \subseteq \Phi(L) \Rightarrow \Phi(\Phi(L)) \subseteq \Phi(\emptyset)$ and similarly, $\Phi(\emptyset) \subseteq L \Rightarrow \Phi(L) \subseteq \Phi(\Phi(\emptyset))$. Combining the two last steps, we conclude that

$$\emptyset \subseteq \Phi(L) \subseteq \Phi(\Phi(\emptyset)) \subseteq \mathcal{L}^{\star} \subseteq \Phi(\Phi(L)) \subseteq \Phi(\emptyset) \subseteq L\,.$$

Extending this logic inductively, each repeated application of the squeezing step produces a tighter bounding pair. Note that this procedure must converge in at most $|L|$ steps since, in each iteration, at least one location ℓ must be added to the lower bounding set or removed from the upper bounding set and there are $|L|$ total locations. A parallel argument, similarly appealing to the monotonicity of Φ, holds when the objective function satisfies SCD-C from below.

The following theorem formalizes the result.

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

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

where k indexes the output of the kth application of the squeezing step. The squeezing step always converges, taking K applications where $K \leq |L|$.

We refer to the iterative application of the squeezing step until convergence as the *squeezing procedure*, and denote the corresponding operator by S^K, so that $\left[\underline{\mathcal{L}}^{(K)}, \overline{\mathcal{L}}^{(K)}\right] = S^K([\emptyset, L])$. If the converged lower and upper bounding sets coincide, then it must be that $\underline{\mathcal{L}}^{(K)} = \mathcal{L}^{\star} = \overline{\mathcal{L}}^{(K)}$, by definition of a bounding pair. If the two bounding sets are not identical, they nevertheless define a (weakly) smaller subdomain of the original CDCP, $\left[\underline{\mathcal{L}}^{(K)}, \overline{\mathcal{L}}^{(K)}\right]$, which we refer to as the *reduced* domain.

The Branching Procedure: Identifying the Optimal Decision Set We introduce a recursive branching procedure as a refinement method that identifies $\mathcal{L}^{\star}$ on the reduced domain. The branching step selects a location ℓ from $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$, creates two "branches" defined by the bounding pairs $\left[\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}\right]$ and $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}\right]$, then applies the squeezing procedure to each. Formally, we define the branching step as follows.

Definition 5 (Branching step)**.** *Given a bounding pair $\left[\underline{\mathcal{L}}, \overline{\mathcal{L}}\right]$ and element $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$, the branching*

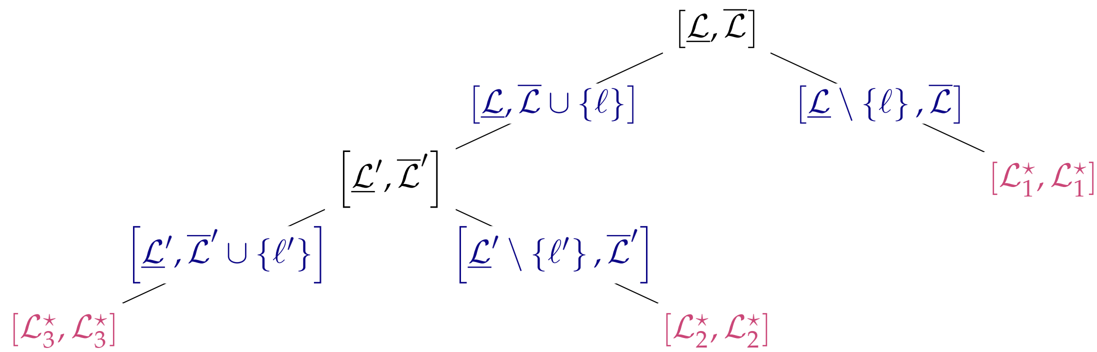


FIGURE 2: AN EXAMPLE PATH OF THE BRANCHING PROCEDURE

This 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, indicated by a terminal node. The final output of the full recursive procedure is the collection of all conditionally optimal sets, in this example $\{\mathcal{L}_1^\star, \mathcal{L}_2^\star, \mathcal{L}_3^\star\}$.

step is defined as

$$B_\ell\left([\underline{\mathcal{L}}, \overline{\mathcal{L}}]\right) = \left\{S^K\left([\underline{\mathcal{L}} \cup \{\ell\}, \overline{\mathcal{L}}]\right), S^K\left([\underline{\mathcal{L}}, \overline{\mathcal{L}} \setminus \{\ell\}]\right)\right\}.$$

Figure 2 illustrates an example of the branching step's application: for a given converged bounding pair, $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, that resulted from the squeezing procedure, two branches are created by applying the branching step with $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$. The branch on the right, which excludes ℓ, converges to a bounding pair with identical lower and upper bounds, identifying $\mathcal{L}_1^\star$ as the optimal decision set *conditional* on excluding ℓ. If squeezing does not converge to identical lower and upper bounds, as on the branch including ℓ on the left, branching recurs. A new location $\ell' \in \overline{\mathcal{L}}' \setminus \underline{\mathcal{L}}'$ is selected among those that remain, creating an additional branch. This recursive process builds a tree, terminating in a conditionally optimal decision set on each branch.

We refer to the recursive application of the branching step until global convergence as the *branching procedure* and define $\Lambda\left([\underline{\mathcal{L}}, \overline{\mathcal{L}}]\right)$ as the collection of conditionally optimal decision sets after global convergence. As we discuss in Section 2.3, the collection of decision sets returned by this recursive procedure turns out not to depend on which locations are selected for branching, so this definition is without ambiguity. The globally optimal decision set is the element in $\Lambda\left([\underline{\mathcal{L}}, \overline{\mathcal{L}}]\right)$ that yields the highest value of the objective function, so that $\mathcal{L}^\star = \arg\max_{\mathcal{L} \in \Lambda([\underline{\mathcal{L}}, \overline{\mathcal{L}}])} f(\mathcal{L})$.

The incumbent alternative to branching is to use the "brute force" method of evaluating the objective function at every element in the reduced domain produced by the squeezing procedure. Intuitively, the branching procedure applies the squeezing procedure as much as possible, yielding to brute force only one location ℓ at a time. As a result, branching is typically faster than brute force. In the worst case, the squeezing step eliminates no decision set on any branch and hence reduces to evaluating the objective function at all elements in the reduced domain just like brute force. In the Online Appendix, we show that this worst case never occurs when the objective function satisfies SCD-C from above.

2.3. The Mathematics of Squeezing and Branching

This section explains the mathematical foundation of our squeezing method and formally connects it to Jia (2008).

Jia (2008), Supermodularity, and SCD-C from Below Jia (2008) introduced the Φ mapping to the economics literature, in the context of a chain store that chooses a set of store locations. The paper recasts finding the optimal decision set of store locations, $\mathcal{L}^\star$, as finding the fixed points of Φ, since $\Phi(\mathcal{L}^\star) = \mathcal{L}^\star$ by construction.[3]

Crucially, Jia (2008) shows that the mapping Φ is order-preserving when the underlying objective function is supermodular, as in the chain store application of the paper. When Φ is order-preserving, the theorem of Tarski (1955) guarantees the existence of well-defined smallest and largest fixed points, $\mathcal{L}^{\inf}$ and $\mathcal{L}^{\sup}$, which together form a natural bounding pair $\left[\mathcal{L}^{\inf}, \mathcal{L}^{\sup}\right]$, since necessarily $\mathcal{L}^{\inf} \subseteq \mathcal{L}^\star \subseteq \mathcal{L}^{\sup}$.

The method of Jia (2008) to identify the bounding pair $\left[\mathcal{L}^{\inf}, \mathcal{L}^{\sup}\right]$ is rooted in Kleene's fixed point theorem, which states that iteratively applying an order-preserving map to $\emptyset$ always converges to its smallest fixed point $\mathcal{L}^{\inf}$, while applying it to L always converges to its largest fixed point $\mathcal{L}^{\sup}$.[4] In cases where the bounding pair $\left[\mathcal{L}^{\inf}, \mathcal{L}^{\sup}\right]$ does not identify $\mathcal{L}^\star$, Jia (2008) applies a brute force search on the reduced domain.

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

Our first contribution relative to Jia (2008) is hence to extend the paper's method beyond the context of supermodular objective functions by showing that SCD-C from below is the

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

[4] See Stoltenberg-Hansen, Lindström, and Griffor (1994) for an exposition. The fixed-point theorem derives from results first established in Kleene (1936) and Kleene (1938)—though it was not stated explicitly there—and is named accordingly.

necessary and sufficient condition for Φ to be order-preserving.

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

This limitation reflects a basic economic feature of negative complementarities. Consider a firm choosing between two perfectly substitutable production locations, $L = \{\text{USA}, \text{CAN}\}$. If no other location is active, either has positive marginal value. However, when both are active, neither does, as each fully substitutes for the other. Applying the algorithm in Jia (2008) yields $\Phi(\emptyset) = \{\text{USA}, \text{CAN}\}$ and $\Phi(\{\text{USA}, \text{CAN}\}) = \emptyset$, so iterating from either extreme oscillates perpetually.

Though repeated application of Φ may not converge to a fixed point with SCD-C from above, it does result in a different type of convergence which alternates between two sets, as in the previous example. We refer to such a pair of sets $[\mathcal{L}, \mathcal{L}']$ for which $\mathcal{L} = \Phi(\mathcal{L}')$ and $\mathcal{L}' = \Phi(\mathcal{L})$ as a "fixed edge" of the Φ mapping. The concept of a fixed edge is a strict generalization of the concept of a fixed point, since if $\mathcal{L}$ is a fixed point of Φ, then the "pair" $[\mathcal{L}, \mathcal{L}]$ is a fixed edge.

Klimeš (1981) shows that, though an order-reversing map need not have a smallest and largest fixed point, it always has a fixed edge $\left[\mathcal{L}^{\text{inf}}, \mathcal{L}^{\text{sup}}\right]$ that is "extreme" in the sense that $\mathcal{L}^{\text{inf}} \subseteq \mathcal{L} \subseteq \mathcal{L}' \subseteq \mathcal{L}^{\text{sup}}$ holds for all other fixed edges $[\mathcal{L}, \mathcal{L}']$.[5] As a result, the extreme fixed edge serves as a bounding pair in the same way the smallest and largest fixed point do with Tarski (1955). With SCD-C from above, the squeezing procedure applied to the trivial bounding pair converges to the extreme fixed edge of Φ.

The insights in this section imply that the squeezing step S is itself an increasing mapping whenever the underlying objective function satisfies either of the SCD-C conditions. As a result, an alternative proof for Theorem 1 follows by applying the theorem of Tarski (1955) and Kleene's theorem to S directly.

The above discussion shows that the squeezing procedure converges to $[\mathcal{L}^\star, \mathcal{L}^\star]$ if and only if $\mathcal{L}^\star$ is the *unique* fixed point of Φ. On the other hand, the branching procedure collects all the fixed points of Φ. In particular, the set of conditionally optimal decision sets at the terminal node of each branch, $\Lambda\left(\underline{\mathcal{L}}, \overline{\mathcal{L}}\right)$, is the collection of all the fixed points of Φ. By implication, the branching procedure results in the same set of conditionally optimal decision sets, regardless of which items are selected for branching, and collapses to brute force only in the extreme case

[5]The intuition behind Klimeš (1981) is simple. Suppose Φ is order-reversing. Then $g \equiv \Phi \circ \Phi$ is order-preserving. By Tarski (1955), g has a set of fixed points with a smallest and largest element. Any fixed point of g must be such that $\Phi(\Phi(\mathcal{L})) = \mathcal{L}$. Then, the pair $[\mathcal{L}, \Phi(\mathcal{L})]$ together must be a fixed edge of Φ.

where every set in the reduced domain returned by squeezing is a fixed point of Φ.

2.4. Solving CDCPs for Heterogeneous Agents

We now extend the combinatorial discrete choice framework to a setting with heterogeneous agents. We consider an objective function $f : \mathcal{P}(L) \times \mathbb{R} \to \mathbb{R}$, that maps a set of items $\mathcal{L}$ and an agent type $z \in \mathbb{R}$ to a scalar payoff $f(\mathcal{L}, z)$. Our exposition focuses on single-dimensional heterogeneity, though our results extend to multidimensional settings (see Arkolakis, Eckert, and Shi 2023).

In the context of heterogeneous agents, the object of interest is the "policy function" that maps an agent's type to its optimal decision set, encoding each agent's solution to the CDCP.

Definition 6 (Policy Function). *The policy function $\mathcal{L}^\star : \mathbb{R} \to \mathcal{P}(L)$ specifies the optimal decision set for each type z, so that $\mathcal{L}^\star(z) = \arg\max_{\mathcal{L} \in \mathcal{P}(L)} f(\mathcal{L}, z)$.*

In the multinational production setting that serves as our example, the policy function maps the productivity of a firm to its optimal combination of production locations. This policy function plays a central role in aggregating firm-level decisions to solve the general equilibrium of the model. For example, computing the aggregate price index requires integrating the policy function over the full support of the firm productivity distribution. We now present an iterative method to recover the policy function for objective functions that satisfy SCD-C and an additional single crossing differences in type condition.

Single Crossing Differences in Type The single crossing differences in type property restricts how a firm's productivity affects the marginal value of a location.

Definition 7 (SCD-T). *The objective function f exhibits single crossing differences in type if, for all items $\ell \in L$, decision sets $\mathcal{L} \in \mathcal{P}(L)$, and types $z, z' \in \mathbb{R}$ such that $z < z'$:*

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

Intuitively, the SCD-T condition ensures that if, as part of a decision set $\mathcal{L}$, a location ℓ has a positive marginal value for a firm of productivity z, then it must also have a positive marginal value for any firm of higher productivity $z' > z$. If instead the marginal value of ℓ crosses zero from above, the problem can be reformulated using the transformation $\tilde{z} \equiv -z$ to satisfy SCD-T.[6]

[6]This property corresponds exactly to the single-crossing condition introduced by Milgrom and Shannon (1994), later referred to as the single-crossing *differences* condition by Milgrom (2004). The firm type z can represent any characteristic, endogenous or exogenous, that affects payoffs, and the policy function $\mathcal{L}^\star(z)$ describes how optimal choices respond to changes in that characteristic. When z is endogenous, the policy function is often referred to as a best-response function.

The squeezing procedure uses the choice monotonicity implied by the SCD-C assumption to rule out many decision sets without explicitly evaluating their payoff. In a similar way, by appealing to the type monotonicity afforded by the SCD-T assumption, our method applies the choice monotonicity implied by the SCD-C restriction to discard many decision sets for entire ranges of productivity without evaluating the objective at any of these productivities. This approach is possible because, when the objective function satisfies both SCD-C and SCD-T, the policy function changes its value only at a finite number of cutoff productivities. As a result, instead of solving a CDCP for every productivity, it suffices to identify the cutoff productivities and the constant value of the policy function in between cutoffs.

When f satisfies supermodularity and SCD-T, Milgrom and Shannon (1994) show that the policy function exhibits a nesting structure: higher productivity firms choose all locations selected by lower productivity firms and possibly more, so that $z < z'$ implies $\mathcal{L}^{\star}(z) \subseteq \mathcal{L}^{\star}(z')$.[7] However, SCD-C and SCD-T alone are not sufficient to ensure nesting.

Solving for the Policy Function with SCD-C and SCD-T To solve for the policy function, we introduce a generalized squeezing procedure. As a first step, we extend the notion of the bounding pair $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, associated with the CDCP of a single firm, to set-valued functions defined over productivities, $\underline{\mathcal{L}}(\cdot)$ and $\overline{\mathcal{L}}(\cdot)$. These "bounding set functions" are such that $\underline{\mathcal{L}}(z) \subseteq \mathcal{L}^{\star}(z) \subseteq \overline{\mathcal{L}}(z)$ for any productivity $z \in \mathbb{R}$. Our solution method iteratively tightens the bounding set functions around the policy function. We refer to $[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)] \equiv [\emptyset, L]$ as the trivial bounding set functions since these constant functions always nest the policy function for all z.

Any pair of bounding set functions implies a partition of productivities into intervals with *identical* lower and upper bounding sets. We define the partition $\mathcal{T}$ created by a pair of bounding set functions $[\underline{\mathcal{L}}(\cdot), \overline{\mathcal{L}}(\cdot)]$ as follows:

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

where t indexes the productivity intervals for which the bounding pair functions imply identical bounding pairs. For brevity, we use $\mathcal{T}$ to denote the partition when the pair of bounding set functions is unambiguous.

Figure 3 illustrates how a pair of bounding set functions partitions the productivity types

[7] More precisely, Milgrom and Shannon (1994) show that nesting holds under the weaker condition of *quasi*-supermodularity, which is stronger than SCD-C as we discuss in the Online Appendix. This result has been widely used in applied theory and empirical work, including Antras, Fort, and Tintelnot (2017) in their analysis of firm-level sourcing decisions. It is also reminiscent of the positive assortative decision patterns studied in Costinot (2009).

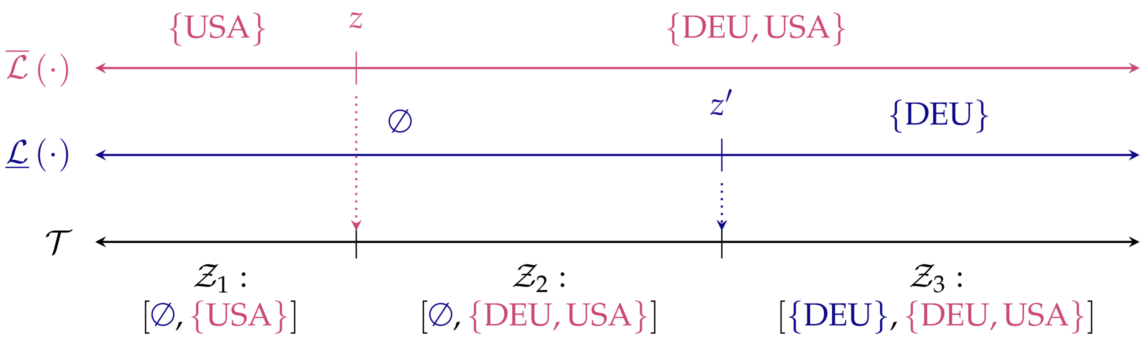


FIGURE 3: PARTITIONING PRODUCTIVITIES BY COMMON BOUNDING PAIRS

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

into intervals. The two lines at the top represent the lower and upper bounding set functions, each changing their values at different cutoffs, z and z'. The line at the bottom shows the resulting partition comprised of productivity ranges for which both bounding set functions are constant. These intervals identify the cutoff productivities at which either the lower or upper bound changes, and thus determine the set of relevant cutoffs for the policy function.

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

Definition 8 (Generalized squeezing step)**.** *Let $\mathcal{C} = \{c : \mathbb{R} \to \mathcal{P}(L)\}$ be the space of functions mapping types to decision sets. The generalized squeezing step is the mapping $S^g : \mathcal{C} \times \mathcal{C} \to \mathcal{C} \times \mathcal{C}$ defined for any $\mathcal{L}(\cdot), \mathcal{L}'(\cdot) \in \mathcal{C}$ as:*

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

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

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

and the functions $z_\ell^g : \mathcal{P}(L) \to \mathbb{R}$ are defined as $z_\ell^g(\mathcal{L}) = \inf\{z \mid D_\ell(\mathcal{L}, z) = 0\}$ for each $\ell \in L$.[8]

When the underlying objective function satisfies SCD-C and SCD-T, the generalized squeezing step applies the logic of the squeezing step from the single-firm case to the full range

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

of productivities. In particular, for each productivity value z, $\Phi^g(\mathcal{L}, z)$ coincides with $\Phi(\mathcal{L})$ when the objective function is evaluated at z.

The importance of the SCD-T assumption is that it eliminates the need to evaluate marginal values separately for each productivity level. Instead, for each location ℓ and decision set $\mathcal{L}$, there exists a *unique* productivity cutoff $z_\ell^g(\mathcal{L})$ that identifies the firm indifferent to including ℓ into $\mathcal{L}$. Rather than computing the marginal value of ℓ at each z, it suffices to check whether a firm's productivity lies above or below the cutoff $z_\ell^g(\mathcal{L})$.

We illustrate how the generalized squeezing step proceeds in practice for an objective function that satisfies SCD-C from above and SCD-T. Consider a pair of bounding set functions and the associated productivity partition $\mathcal{T}$. For a given interval $\mathcal{Z}_t \in \mathcal{T}$, we compute the two cutoffs $z_\ell^g(\underline{\mathcal{L}}_t)$ and $z_\ell^g(\overline{\mathcal{L}}_t)$ for each ℓ, where the SCD-C and SCD-T conditions together imply $z_\ell^g(\underline{\mathcal{L}}_t) \leq z_\ell^g(\overline{\mathcal{L}}_t)$. Then, for all firms with productivity $z < z_\ell^g(\underline{\mathcal{L}}_t)$ in $\mathcal{Z}_t$, ℓ is not part of the optimal decision set, so the upper bounding set function updates to $\overline{\mathcal{L}}_t \setminus \{\ell\}$ for these productivities. Conversely, for all firms with productivity $z_\ell^g(\overline{\mathcal{L}}_t) < z$ in $\mathcal{Z}_t$, ℓ is part of the decision set, so the lower bounding set function updates to $\underline{\mathcal{L}}_t \cup \{\ell\}$ for these productivities. Figure 3 depicts the outcome of updating the trivial bounding set functions where z corresponds to $z_\ell^g(\underline{\mathcal{L}}_t)$, z' corresponds to $z_\ell^g(\overline{\mathcal{L}}_t)$, and ℓ corresponds to DEU. A full application of the generalized squeezing step requires computing $2 \times |L|$ cutoffs for each interval $\mathcal{Z}_t$.

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

Theorem 2. *If the objective function f exhibits SCD-C and SCD-T, iteratively applying the generalized squeezing step to the trivial bounding set functions $\left[\underline{\mathcal{L}}^{(0)}(\cdot), \overline{\mathcal{L}}^{(0)}(\cdot)\right] = [\emptyset, L]$ yields a sequence of bounding set functions so that for all $z \in \mathbb{R}$,*

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

where k indexes the output of the kth application of the generalized squeezing step. The generalized squeezing step always converges, taking K' applications where $K' \leq |L|$.

At convergence, the optimal decision set is identified for any interval $\mathcal{Z}_t^{(K')} \in \mathcal{T}^{(K')}$ where the bounds coincide, that is, $\overline{\mathcal{L}}_t^{(K')} = \underline{\mathcal{L}}_t^{(K')} = \mathcal{L}_t^{\star}$. In intervals where the bounds differ, the generalized squeezing step has converged without identifying the optimal decision set. For such cases, we define two refinement methods that always converge to the policy function, including a *generalized branching* procedure, in the Online Appendix.

In what follows, we refer to the application until convergence of generalized squeezing and its refinement as the "policy function method." In contrast to discretization or bisection methods, the policy function method uses the monotonicity afforded by the SCD-T restriction to identify the exact cutoffs at which the policy function changes its value. This approach eliminates unnecessary computation at non-cutoff productivities. By recovering the exact policy function, we also avoid the interpolation between productivities required by discretization methods. Section 5.1 shows that such interpolation can introduce substantial errors in aggregate variables in the context of a realistically calibrated model of multinational production.

3. A Quantitative Model of Multinational Production

In this section, we introduce a quantitative model of multinational production (MP). We characterize the firm's location problem and the economic mechanisms that turn it into a CDCP, then derive explicit parameter conditions under which SCD-C and SCD-T hold. In the Online Appendix, we extend the framework to accommodate a broader class of demand and cost functions.

Setup The world economy consists of a discrete set of countries L. We index firm headquarter locations by i, production locations by ℓ, and final consumption locations by n. Each firm produces a differentiated final good variety ω. Every production location ℓ has a mass of households H_ℓ, each of which inelastically supply one unit of labor at wage w_ℓ. Labor markets are perfectly competitive and output markets are monopolistically competitive.

Demand System Consumers in all destinations have identical CES preferences with elasticity σ over the set of available final goods. Then, the demand function for good ω as a function of its destination-specific price $p_n(\omega)$ is

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

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

Production Technology Consider a firm headquartered in location i that produces the final good ω and operates a set of production locations $\mathcal{L} \subseteq L$. The firm has productivity $z(\omega)$, which describes its efficiency of producing good ω in any potential location ℓ. We index firms by i and z alone, anticipating that firms with identical headquarter location and productivity behave identically.

The firm delivers its final good to destination n by combining production from its locations ℓ. The resulting marginal cost of a firm headquartered in location i of delivering a unit of its output to destination n is given by a constant-elasticity aggregator over the marginal cost of each of its active production locations:

$$c_{in}\left(\mathcal{L},z\right)=\frac{1}{z}\left[\sum_{\ell\in\mathcal{L}}\xi_{i\ell n}^{-\theta}\right]^{-\frac{1}{\theta}}\qquad\text{where}\qquad\xi_{i\ell n}=\gamma_{i\ell}\frac{w_\ell}{T_\ell}\tau_{\ell n}.\tag{1}$$

For each production location, w_ℓ is the equilibrium wage rate and T_ℓ is an exogenous location-specific productivity shifter common to all firms producing there. Firms also face a bilateral cost of multinational production $\gamma_{i\ell}$, or MP cost, which captures factors such as communication or language costs, as well as a bilateral cost of trade $\tau_{\ell n}$. We summarize all cost shifters in a trilateral resistance term denoted $\xi_{i\ell n}$.

All else equal, the marginal cost in equation (1) increases in wages, MP costs, and trade costs, while it decreases in firm and country productivity. Crucially, the marginal cost always declines as the production location set $\mathcal{L}$ grows: a firm that operates an expanded set of production sites has lower marginal cost of supplying its good to any final destination.

The elasticity of substitution across locations, $\theta > 0$, captures that the firm may require all of its locations for production. In the limiting case as $\theta \to \infty$, the firm uses only its lowest-cost location.

The elasticity θ determines whether there are increasing, constant, or decreasing returns to additional locations on the cost side. For $\theta > 1$, locations are substitutes so that, all else equal, any additional production location lowers marginal costs by less; on the other hand, with $0 < \theta < 1$, locations are complements so that any additional production location lowers marginal costs by more.

Several microfoundations can give rise to the cost aggregator in equation (1). In Appendix A, we present one based on Fréchet-distributed cost shocks, closely aligned with the frameworks in Antras, Fort, and Tintelnot (2017) and Tintelnot (2017). In Antras, Fort, and Tintelnot (2017), final-good firms do not operate sourcing locations themselves or export; instead, they pay fixed costs to add sourcing partners, each subject to Fréchet-distributed cost shock draws. The setup in Tintelnot (2017), which is closer to ours, features firms that produce and export bundles of goods and pay fixed costs to establish plants across locations. The marginal cost of producing each good at a plant depends on plant-specific Fréchet shocks. While these models differ in structure, they ultimately deliver the same cost aggregator. Importantly, conditional on equation (1), the equilibrium of our model implies the same aggregate allocations regardless of the specific microfoundation.[9]

Profit Maximization The profit maximization problem of the firm has two parts: choosing a set of production locations $\mathcal{L}$ and setting the price of its final good in each destination market. Firms set the price for their final good in destination n as a constant markup over their marginal cost, $c_{in}(\mathcal{L},z)$, so that

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

Finally, we turn to the choice of the production location set $\mathcal{L}$. To produce in location ℓ, a firm headquartered in i must pay a fixed cost $f_{i\ell}$, denominated in location ℓ production labor. Without the fixed cost $f_{i\ell}$, all firms would produce in all locations ℓ and there would be no (combinatorial) discrete choice problem to solve.

The firm headquartered in location i with productivity z chooses the optimal location decision set $\mathcal{L}^{\star}(z)$ to maximize its operating profits:

$$\max_{\mathcal{L}\in\mathcal{P}(L)} \pi_i(\mathcal{L},z) \equiv \max_{\mathcal{L}\in\mathcal{P}(L)} \left\{ \sum_n \frac{1}{\sigma-1} q_{in}(\mathcal{L},z)\, c_{in}(\mathcal{L},z) - \sum_{\ell\in\mathcal{L}} w_\ell f_{i\ell} \right\}, \tag{2}$$

where $q_{in}(\mathcal{L},z) \equiv q_n(p_{in}(\mathcal{L},z))$ is the demand for the final good in destination market n.

To establish a headquarter in location i, firms first must pay a labor-denominated entry cost f_i^e to draw a productivity z from an exogenous distribution $G_i(z)$. Given their productivity z, firms with non-negative operating profits produce positive quantities, while firms with negative operating profits shut down right after paying f_i^e.

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

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

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

The second equilibrium condition is a free entry condition that reflects that firms enter until their expected operating profits before drawing a productivity are zero. The free entry

[9] In particular, Tintelnot (2017) assumes that the elasticity of substitution within and across firms is identical, which requires the Fréchet shape parameter to exceed this common elasticity, restricting the problem to the SCD-C from above case as we discuss below. By allowing the two elasticities to differ, as in our setup, this restriction is relaxed. The Fréchet distribution can also be replaced by a multivariate Pareto, as in Arkolakis, Ramondo, et al. (2018). Alternatively, one can microfound the cost aggregator by assuming that firms produce intermediate goods in each location that are incomplete substitutes, in which case the Fréchet parameter is replaced by the elasticity of substitution across intermediates, as in Antràs et al. (2024a).

condition pins down the total mass of entrants M_i in each origin country i by equalizing the entry cost on the left with the operating profits of the average firm across all destination markets on the right:

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

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

$$P_n^{1-\sigma} = \sum_i M_i \int_{\tilde{z}_i}^{\infty} p_{in} \left(\mathcal{L}_i^\star \left(z \right), z \right)^{1-\sigma} \mathrm{d}G_i \left(z \right). \tag{5}$$

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

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

where the indicator $\mathbb{1}_{i\ell}\left(z\right) \equiv \mathbb{1}\left\{\ell \in \mathcal{L}_i^\star\left(z\right)\right\}$ takes the value 1 if a firm with productivity z from origin i opens a production location in country ℓ.[10] The term X_n denotes total expenditure on final goods in destination n.

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

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

The general equilibrium in our model is defined as follows.

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

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

Before describing how to quantify the model using data on trade and MP flows, we first describe how to establish the SCD-C and SCD-T properties in our framework.

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

Establishing the SCD-C and SCD-T conditions Our model features two forces that generate interdependencies among the production locations of an individual firm. On the demand side, the demand elasticity σ indexes the firm's ability to scale. The larger is σ, the more sensitive demand is to prices, and hence the more the marginal cost savings of large multinationals translates into larger sales. All else equal, such returns to scale make each additional location more valuable, generating positive complementarities among locations. On the cost side, when $\theta > 1$, production locations are substitutes as they compete with one another to supply any destination ("cannibalization"). However, if $0 < \theta < 1$, locations act as complements in production. Overall, when $\theta > 1$, the cost- and demand-side effects oppose each other; when $\theta < 1$, they reinforce each other.

The model generates a closed-form condition on parameters that determines the type of complementarities among locations in the firm's CDCP. As we show formally in the Online Appendix, negative complementarities dominate and the firm's profit function satisfies SCD-C from above iff $\sigma - 1 < \theta$, whereas positive complementarities prevail and SCD-C from below holds iff $\sigma - 1 > \theta$.

To build intuition for these parameter conditions, consider a symmetric version of our economy where $\xi_{i\ell n} = \xi$. In this economy, the marginal value of location ℓ is given by:

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

where $\mathcal{G}$ is a composite general equilibrium constant and, since all locations are symmetric, only the *number* of locations matters, not their precise identity. The cost and demand elasticities interact to determine the strength and direction of the complementarities among production locations. Equation (8) shows that if $\sigma - 1 > \theta$, the marginal value of any given production location ℓ is increasing in the number of other locations, so that SCD-C from below holds, indicating positive complementarities. If instead $\sigma - 1 < \theta$, complementarities are negative and SCD-C from above holds. In the special case where $\sigma - 1 = \theta$, the marginal value of ℓ is independent of the firm's decision set $\mathcal{L}$.[11] Finally, SCD-T requires that a location's marginal value is higher at more productive firms, which is the case as long as $\sigma > 1$.

Equation (8) highlights that firms in our model face a CDCP because of the combination

[11] In the Online Appendix, we also show how to verify SCD-C and SCD-T for more general marginal cost functions $c_{in}(\mathcal{L}, z)$ of which the cost function in this section is a special case. Our generalized framework nests production structures as in Ramondo (2014), Arkolakis, Ramondo, et al. (2018), Lind and Ramondo (2023), and Xiang (2022), that is, production with correlated idiosyncratic shocks across locations, or with multiple different technological methods of production. In general, regardless of the cost function, the strength of the scale complementarities depends on how changes in marginal costs translate into differences in variable profits, which is determined by the elasticity of demand and the pass-through elasticity of costs to price. In the present model, markups are constant so that the pass-through elasticity is always 1; consequently, only the elasticity of demand matters.

of two ingredients: complementarities among locations and the fixed costs of setting up production locations. Without complementarities, that is when $\theta = \sigma - 1$, the marginal value of any production location is independent of all others and hence the firm faces $|L|$ independent decisions. Without fixed costs, that is, when $f_{i\ell} = 0 \, \forall i, \ell$, the marginal value of all locations is always positive so that all firms establish production locations in all countries, regardless of complementarities.

4. Quantification

In this section, we provide an overview of our model calibration, which follows standard techniques and borrows central parameters from the literature. For use in our quantitative exercises in Section 5, we calibrate the model twice, once with negative and once with positive complementarities. We relegate most details of the quantification, such as of the treatment of the data and the model's fit, to Appendix C.

4.1. Data

We obtain information on manufacturing trade and multinational production for 32 countries from Alviarez (2019). For each host country, the data set contains the number of manufacturing enterprises and their total manufacturing sales by origin country. For example, the data contain the number of German manufacturing enterprises in France and their total sales. The dataset also contains bilateral trade flows for all country pairs. For all variables, the data reflect average values over the period from 2003 to 2012.[12]

While the dataset contains foreign affiliate counts for each nation, it lacks each country's total enterprise count and information on firm entry and survival. To supplement the dataset, we obtain counts of total enterprises for each country from the OECD Structural Statistics of Industry and Services dataset, and data on the 1-year survival rates of enterprises in each country from the OECD Structural and Demographic Business Statistics.

To construct standard bilateral gravity controls, we use the CEPII database (see Conte, Cotterlaz, Mayer, et al. 2023), which provides measures of geographic distance, a shared border indicator, a colonial ties dummy, and a common language dummy. In addition, we use the TRAINS dataset to construct bilateral manufacturing tariff measures, following standard practices in the literature. Data on real GDP per capita and total employment are drawn from

[12]To construct the dataset, Alviarez (2019) combined data from the OECD, the Eurostat Foreign Affiliate Statistics database, the Bureau of Economic Analysis, and Bureau van Dijk's Orbis dataset. The dataset contains information on the following countries: Australia, Austria, Belgium, Bulgaria, Canada, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Japan, Latvia, Lithuania, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Romania, the Russian Federation, Slovakia, Spain, Sweden, Turkey, Ukraine, the United Kingdom, and the United States.

Parameter	Description	Value/Target Moment
Externally Calibrated		
σ	Demand elasticity	Set to 4 (Arkolakis, Ramondo, et al. 2018)
$\frac{\sigma-1}{\theta}$	Location complementarity	Set to $\frac{2}{3}$ (negative complementarities) and $\frac{3}{2}$ (positive complementarities)
Internally Calibrated		
ζ	Firm Pareto shape	Sales distribution right-tail (Arkolakis 2010)
$\underline{z}_i$	Firm Pareto minimum	Total foreign MP (Alviarez 2019)
T_ℓ	Location productivity	GDP per capita (PWT)
f_i	Fixed cost (origin component)	1-year firm survival rate (OECD)
f_i^e	Entry cost	Total enterprise counts (OECD)
H_ℓ	Labor supply	Employment (PWT)
$\tau_{\ell n}, \gamma_{i\ell}, \nu_{i\ell}$	Bilateral trade, MP, fixed costs	Gravity coefficients on trade flows, MP flows, affiliate counts (Appendix C.2)

Table 1: Parameters and Target Moments

This table summarizes the calibration strategy of the model. Two parameters, σ and θ, are externally calibrated; the remaining are calibrated simultaneously in a single iterative routine that targets the moments described in the last column. For each empirical moment, we list the data source. PWT refers to the Penn World Tables and OECD to either the OECD Structural Statistics of Industry and Services dataset or for the 1-year survival rates of enterprises in the OECD Structural and Demographic Business Statistics.

the Penn World Tables (Feenstra, Inklaar, and Timmer 2015).

4.2. Calibration Strategy

Our calibration strategy uses indirect inference to identify the model's internally calibrated elasticities. For a given set of elasticity values, we choose productivity parameters to match country-specific aggregates in general equilibrium. Once these targets are met, we use the model's equilibrium allocations to estimate a set of aggregate gravity regressions and update the elasticity guesses to match the corresponding coefficients in the data. Since matching country-level aggregates in general equilibrium is itself an iterative process that requires repeatedly solving for the policy function, our method is essential to the feasibility of this approach. Table 1 summarizes the moments targeted by each parameter. Further details are in the Online Appendix.

Demand Elasticity and Location Substitution Elasticity: σ and θ Together, the elasticity of substitution σ across consumption goods and the elasticity of substitution θ across production locations in the firm's cost function determine the type of SCD-C condition satisfied by the firm location problem. In our baseline calibration, we follow Arkolakis, Ramondo, et al. (2018)

and set $\sigma = 4$.[13]

We calibrate the model twice, with two different values for θ, to conduct quantitative exercises with both negative and positive complementarities. For the negative complementarity calibration, we follow Arkolakis, Ramondo, et al. (2018) and set $\theta = 4.5$. Combined with the value for σ, the degree of complementarity is $\frac{\sigma-1}{\theta} = \frac{2}{3}$. For symmetry, we set θ in the positive complementarity calibration so that $\frac{\sigma-1}{\theta} = \frac{3}{2}$.

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

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

For the calibration, we decompose the fixed cost of setting up a production location in ℓ for firms headquartered in location i into an origin-specific shifter and a bilateral term, $f_{i\ell} \equiv f_i \nu_{i\ell}$. For a given entry cost f_i^e, we choose the base component of fixed costs f_i to match the share of "surviving" firms whose productivity draw exceeds the cutoff level defined by the zero profit condition in equation (3). In a second step, we recover the entry cost f_i^e by inverting the free entry condition in equation (4) so that the model matches the implied number of entering firms observed in the data, which we calculate as the number of enterprises divided by the survival rate. We set total labor supply in each location, H_i, to match total employment in each country.

Trade Costs, MP Costs, Fixed Costs In the theory, three bilateral cost matrices shape the patterns of trade flows, foreign affiliate sales, and foreign affiliate counts across country pairs: the matrix of trade costs $\{\tau_{\ell n}\}_{\ell n}$, the matrix of MP costs $\{\gamma_{i\ell}\}_{i\ell}$, and the matrix of the bilateral components of fixed costs $\{\nu_{i\ell}\}_{i\ell}$. To calibrate these parameters, we use data on trade flows, MP flows, and bilateral affiliate counts as indicated in Table 1.

[13]Our choice of σ falls within the range of estimates of Broda and Weinstein (2006). Arkolakis, Ramondo, et al. (2018) show that it is also consistent with markup estimates from the manufacturing sector. Moreover, our value is similar to the estimate of $\sigma = 3.89$ from Head and Mayer (2019). While Head and Mayer (2019) focuses on multinational production in the car industry, their estimation strategy is consistent with our theoretical setup and also generates markups in line with the microeconomic estimates from the car industry in Goldberg (1995) and Berry, Levinsohn, and Pakes (1995). Finally, our choices of θ generates a trade elasticity of 2.7. Head and Mayer (2014) report a median estimate of the trade elasticity of 3.19 and a mean of 4.51 based on a meta-analysis of 32 papers.

We follow Tintelnot (2017) and parameterize the three matrices $\{\tau_{\ell n}, \gamma_{i\ell}, \nu_{i\ell}\}$ as constant elasticity functions of the standard bilateral gravity variables: geographic distance, a shared colonial past indicator, a shared language indicator, and a common border indicator. We specify a separate elasticity for every gravity variable in each of the three matrices, leading to a total of twelve elasticities to estimate. We present the expressions for each element of the bilateral cost matrices in Appendix C.2.[14] We then choose the elasticities of the various gravity variables in the bilateral cost matrices using an indirect inference procedure. We specify three gravity equations, for the trade flows, MP flows, and bilateral affiliate counts both in model and data and choose the elasticities of the gravity variables in the model to match the regression coefficients found in the data.

5. Quantitative Exercises and Counterfactuals

In this section, we use our quantified model to conduct a number of numerical exercises illustrating the performance of our solution method. We also use our model to measure the welfare gains from multinational production, in particular analyzing how they differ under negative or positive complementarities and in settings with or without fixed costs.

5.1. Computational Performance

We conduct three different numerical experiments, demonstrating advantages of our solution method in terms of computational speed, precision, and breadth of applicability. For experiments that vary the number of countries, we sample from the the cost and productivity parameters of our calibrated model to generate synthetic countries.[15]

Speed We compare the speed of our policy function method ("Policy") against two alternative discretization approaches on a grid of $2^{14} \approx 16000$ grid points: the "naive" approach which uses brute force to solve the CDCP at each grid point by evaluating profits for all possible location combinations and the "squeezing" approach which applies the squeezing procedure at each grid point until convergence, then brute force to the reduced domain. Unlike our "policy" method, these two grid-based approaches generate discretization error in the computed policy functions.

Table 2 reports the time (in seconds) required to compute the policy function across varying numbers of synthetic countries. We compute $\mathcal{L}_i^\star(z)$ for firms from each origin i and report the average time across origins. The naive method requires more than an hour even with

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

[15] A synthetic country is described by fundamentals $\{T_\ell, f_i\}$, aggregates $\{w_\ell, P_\ell\}$, and bilateral costs $\{\tau_{\ell n}, \gamma_{i\ell}, \nu_{i\ell}\}$. For each of these objects, we non-parametrically fit a distribution to the estimated values from which we then sample to generate synthetic countries.

Countries	Negative Complementarities			Positive Complementarities		
	Naive (1)	Squeezing (2)	Policy (3)	Naive (1)	Squeezing (2)	Policy (3)
8	8	0.423	0.019	7	0.480	0.034
16	5454	2.261	0.039	4356	2.364	0.087
32	–	11.093	0.107	–	13.217	0.186
64	–	66.001	1.323	–	94.459	1.293
128	–	456	14.061	–	795	14.702
256	–	3239	331	–	6479	374
Grid points	2^{14}	2^{14}	–	2^{14}	2^{14}	–

TABLE 2: RUNTIMES FOR DIFFERENT SOLUTION METHODS

This table illustrates the computational time in seconds for computing the firm policy function by discretizing firm productivity, then solving the firm's CDCP at each grid point by (1) evaluating the profit function for all possible production location combinations ("Naive"); or (2) applying squeezing, then brute force ("Squeezing"). We contrast these methods with our policy function method ("Policy") which does not require discretization. We repeat the exercise in both the calibration with negative and positive complementarities. Synthetic countries are generated by sampling from the distribution of parameter estimates obtained from the calibrated model. We then time how long it takes to solve for the policy functions $\mathcal{L}_i^\star(\cdot)$ for all origins, and report the average time taken across origins. Trials were computed on an Apple M1 (2020) CPU.

just 16 countries, while the policy method solves the same problem in less than a tenth of a second. Even for 128 countries, the policy method recovers the average policy function in under 15 seconds, while it takes about 6 minutes to compute the average policy function with 256 countries.

In the case of positive complementarities, the squeezing method corresponds to the incumbent approach pioneered by Antras, Fort, and Tintelnot (2017), and our policy function method is roughly an order of magnitude faster across all country counts. For negative complementarities, the naive brute-force approach represents the benchmark, and our method improves on it by four orders of magnitude. Approximately three-quarters of the speed gains come from extending the squeezing method of Jia (2008) to the negative complementarity case, and the remaining quarter from introducing our policy function method, which avoids solving the CDCP at every grid point.

Precision Table 2 understates the performance gap between the policy method and the alternative approaches, as it does not account for the discretization error inherent in the latter. Discretization error arises because outcomes must be interpolated between grid points, so that selecting the number of grid points involves a trade-off between computational time and discretization error.

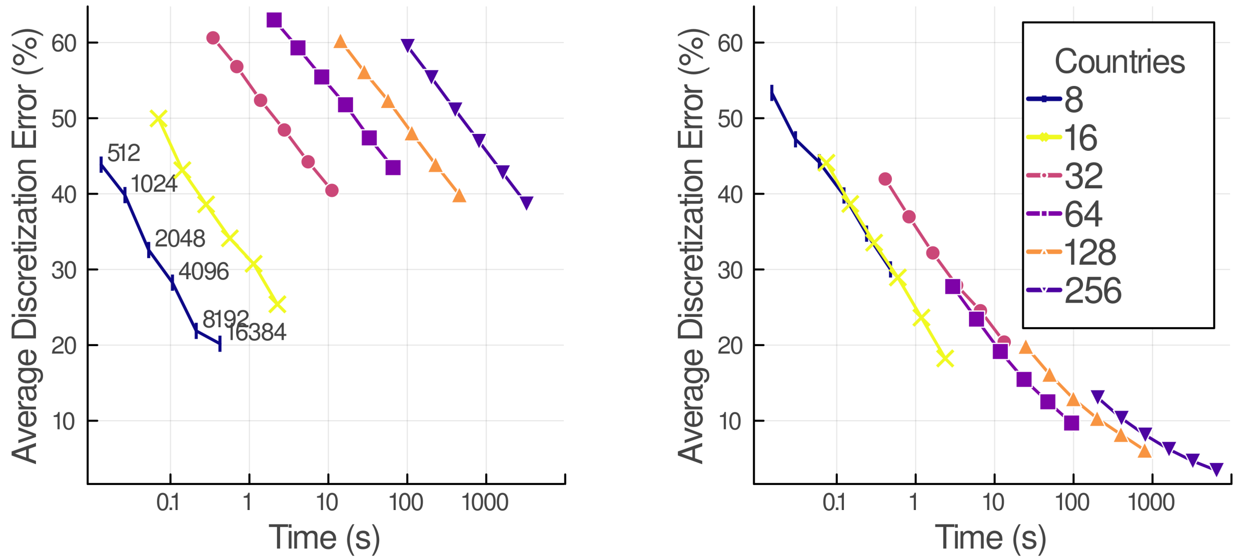
Our policy function method solves for the *exact* policy function, allowing us to quantify the discretization error of other methods. To do so, we first compute the total value of trade flows from origin i to destination n via production in country ℓ, denoted $X_{i\ell n}$, by aggregating over the exact policy function. Then, we compute the corresponding approximation $\hat{X}_{i\ell n}$ by aggregating over the discretized policy function of the "Squeezing" method, interpolating between grid points. We define the discretization error as the average percentage deviation of flows across all country triplets, that is, $N^{-3}\sum_{i,\ell,n}\left|\hat{X}_{i\ell n}/X_{i\ell n}-1\right|\times 100\%$.

Figure 4 graphs the discretization error against computational time for varying grid densities and both types of complementarities. Each line shows the precision-time frontier for a different number of countries. With 512 grid points and 8 countries, computation is fast, but the average discretization error exceeds 40% regardless of the type of complementarity. Even with over 16000 grid points, the error remains above 20% in both cases. Doubling the number of grid points reduces the error by 5 to 10 percentage points but doubles the computation time. As the number of countries increases, the frontier shifts rightward at a log-constant rate, reflecting the polynomial complexity of the generalized squeezing procedure. With negative complementarities, average discretization error increases in the number of countries. By contrast, it decreases with positive complementarities, since the nesting structure of the policy function mitigates discretization error.

Wide Applicability In a final exercise, we examine how computational time depends on the strength and direction of complementarities, summarized by the ratio $(\sigma-1)/\theta$. We recalibrate the 32-country model (following Section 4) for values of θ ranging from 0.8 to 20, yielding $(\sigma-1)/\theta$ values between 0.15 and 3.9. This range extends substantially beyond the strength of complementarities suggested by estimates of θ and σ in the literature and notably also includes cases where $0<\theta<1$, so that production locations are complements in the firm's cost function.

The left panel of Figure 5 shows the average time required to solve for the policy function across origin countries, using our policy function method for every level of complementarity. The figure also shows the fraction of the total computational time spent using the squeezing method. Across the range of complementarities, our policy function method takes under 0.26 seconds. Computation is fastest near the vertical line, which marks the special case of no complementarity, where production locations are effectively independent. As the degree of complementarity increases in either direction, computation time increases but never more than doubles.

Our theoretical results help explain the patterns in the left panel. Theorem 2 establishes that the generalized squeezing method never takes more than $|L|$ applications, consistent with little



(A) Negative Complementarities

(B) Positive Complementarities

FIGURE 4: THE PRECISION-TIME FRONTIER WITH DISCRETIZATION METHODS

This figure illustrates the discretization error in computing the policy function with the "Squeezing" approach, for different numbers of countries. To measure discretization error, we compute the percentage deviation for each trilateral flow $X_{i\ell n}$, when using the discretized policy function compared to the true policy function. In particular, we compute the average discretization error as $N^{-3} \sum_{i,\ell,n} \left| \hat{X}_{i\ell n} / X_{i\ell n} - 1 \right| \times 100\%$, where $X_{i\ell n}$ are the sales of location ℓ production sites, whose headquarters are in location i, to destination market n computed using the "Policy" method (that is, the exactly solved model) and $\hat{X}_{i\ell n}$ is the same object in the model computed using the "Squeezing" method. Trials were computed on an Apple M1 (2020) CPU.

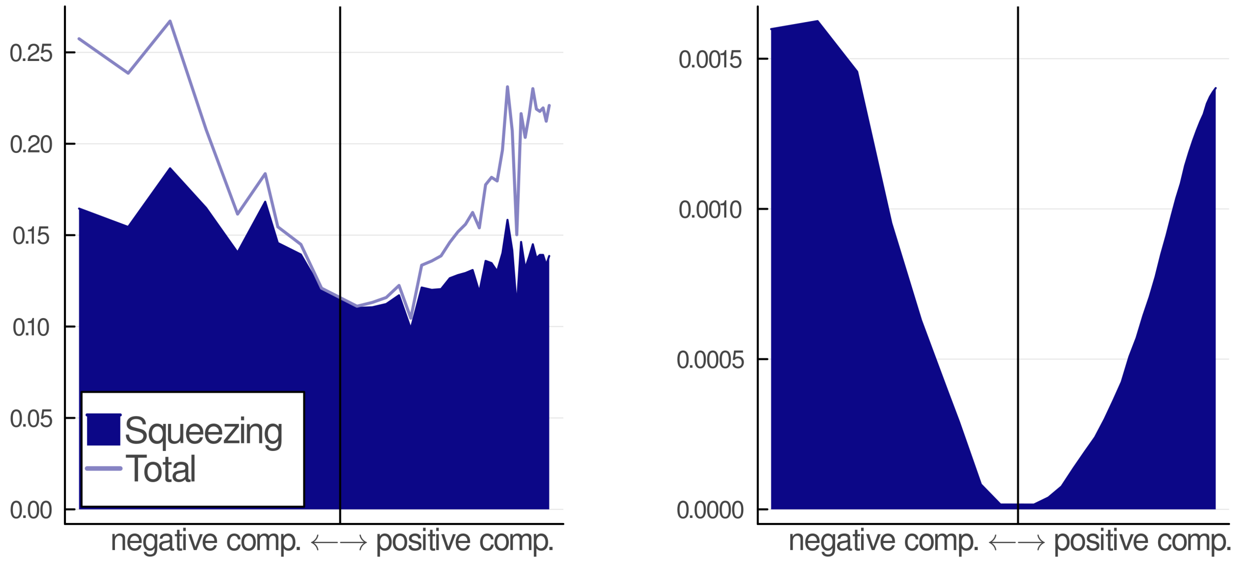
variation in squeezing time across the full range of complementarities.

The right panel shows a proxy measure for the effectiveness of the squeezing procedure: the number of locations which separate the upper and lower bounding set averaged across all firms and countries after convergence of the generalized squeezing procedure. In general, there are very few locations separating the bounding sets after convergence. At the same time, the average number of leftover locations increases with stronger complementarities, consistent with the increasing gap between squeezing and total time in the left panel as complementarities grow stronger.

5.2. CDCPs and the Welfare Gains from Multinational Production

Complementarities among production locations, together with the fixed costs of setting them up, give rise to the CDCP in our model. An alternative modeling approach is to abstract from either complementarities or fixed costs to avoid solving CDCPs. In this section, we examine how the welfare gains from multinational production depend on the presence of both forces.

To quantify the role of these assumption for the welfare gains from multinational production,



(A) Compute Time (s)

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

FIGURE 5: COMPLEMENTARITIES AND COMPUTATIONAL PERFORMANCE

This figure shows the performance of the policy function method with different degrees of complementarity. The strength of complementarity is measured by the ratio $\frac{\sigma-1}{\theta} / \left(1 + \frac{\sigma-1}{\theta}\right)$, and varied by changing θ while holding $\sigma = 4$ fixed (its value in the baseline calibration). The vertical line indicates the case with no complementarities, with $\frac{\sigma-1}{\theta} = 1$. The panel on the left shows the total policy function method computation time in seconds as well as the time spent on generalized squeezing in particular. The right panel shows the number of locations which separate the upper and lower bounding set, averaged across all firms and countries, after convergence of the generalized squeezing procedure. Trials were computed on an Apple M1 (2020) CPU.

we extend the welfare formula in Arkolakis, Costinot, and Rodríguez-Clare (2012) to the context of our model with both complementarities and fixed costs.[16] The following equation describes the welfare impact in country i of imposing trade or MP autarky by setting the respective bilateral costs to infinity:

$$\ln \frac{\hat{w}_i}{\hat{P}_i} = \underbrace{\ln \hat{\pi}_{iii}^{-\frac{1}{\sigma-1}}}_{\text{openness}} + \underbrace{\ln \hat{M}_i^{\frac{1}{\sigma-1}} + \ln \hat{\tilde{z}}_i^{-\frac{\zeta}{\sigma-1}}}_{\text{varieties}} + \underbrace{\ln \hat{\tilde{z}}_i + \ln \left[\sum_{\mathcal{Z}_i^t \in \mathcal{T}_i} \lambda_{iii}^t \left(s_{iii}^t\right)^{\frac{\sigma-1}{\theta}-1} \right]^{\frac{1}{\sigma-1}}}_{\text{average productivity}} \tag{9}$$

where $\hat{x} = x'/x$ and x denotes the value of a variable in our baseline calibration and x' its value under autarky. We denote by π_{iii} the fraction of all final spending in i on goods produced

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

in i by firms headquartered in i. This own share is an inverse measure of openness of i to *both* trade and multinational activity. M_i is the mass of entrants in location i and $\tilde{z}_i$ is the survival productivity cutoff in location i for firms headquartered in i. The policy function $\mathcal{L}_i^{\star}(\cdot)$ induces a partitioning $\mathcal{T}_i = \left\{\mathcal{Z}_i^t\right\}_t$, so that $\mathcal{L}_i^{\star}(z) = \mathcal{L}_i^t$ for all $z \in \mathcal{Z}_i^t$. Out of all sales in i by firms from i in interval t, the term s_{iii}^t denotes the share that is produced in i so that $s_{i\ell i}^t$ sums to 1 across production locations ℓ, within each interval t. Out of total sales in i by firms from i produced in i, λ_{iii}^t denotes the share accounted for by firms in interval t, so that λ_{iii}^t sums to 1 across all intervals t.

Equation (9) shows that the welfare effects of moving to trade or MP autarky work through three channels: first, a standard *openness channel* that captures a reduction in real consumption; second, a *varieties channel* that adjusts for changes in the number of domestic varieties; and third, an *average productivity channel* that adjusts for changes in the average productivity with which domestic goods are produced.

Since trade and MP autarky shift inward the effective production possibility frontier of all countries, the openness channel is typically negative. The signs of the variety and productivity channels, however, tend to depend on the degree to which countries are headquarter locations for multinationals compared to host countries for the foreign affiliates of multinationals headquartered abroad.

The variety effect reflects that both trade and MP autarky shrink the profits of previously large firms engaged in these foreign activities. The resulting reduction in local wages allows the selection cutoff to fall and more local firms to survive, creating new varieties that benefit welfare. Relative to the no-complementarity case, firms engaged in MP are on average larger and more profitable in the positive complementarity case than in the negative complementarity case, since complementarities directly shape both the marginal cost advantage attainable through MP and the market size advantage attainable through trade. As a result, in the case of positive complementarities, the variety effect is more positive compared to the case of negative complementarities.

The average productivity channel captures both extensive and intensive margin effects of the move to autarky. On the extensive margin, average firm productivity declines because the entry threshold shifts downward ($\hat{\tilde{z}}_i < 1$), allowing lower-productivity firms to survive. On the intensive margin, captured by the last term in the expression, the sales-weighted average productivity of operating firms changes as relative firm sizes adjust. When locations are complements ($\sigma - 1 > \theta$), the most productive firms in each origin country i shrink the most, as they lose the scale economies that previously amplified their productivity advantage and supported their large domestic market share. As a result, reallocation among incumbents in the domestic market implies that the sales-weighted average productivity declines and

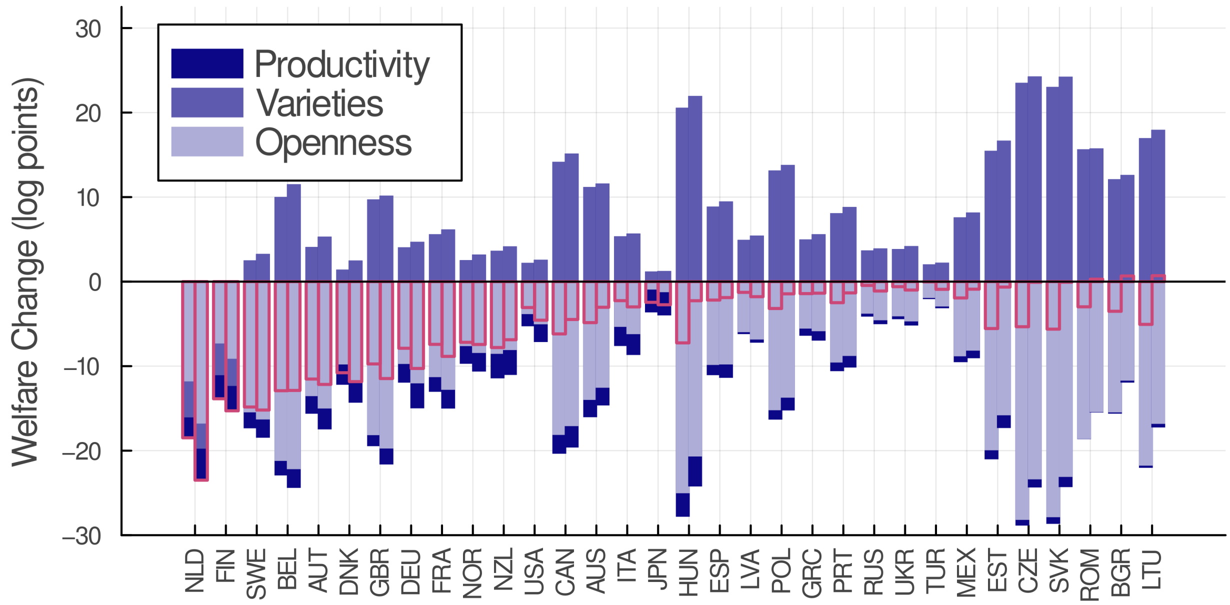


FIGURE 6: WELFARE CONSEQUENCES OF IMPOSING MP AUTARKY

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

the intensive margin effect is always negative. By contrast, when locations are substitutes ($\sigma - 1 < \theta$), the most productive firms expand relative to other firms in the domestic market, as they substitute foreign production with domestic production. Thus, the sales-weighted average productivity increases and the intensive margin effect is always positive.

We now use the two calibrations of our model to study the welfare consequences of moving to MP autarky with negative versus positive complementarities. We focus on MP autarky because it highlights the role of multinational location choices with complementarities and fixed costs, that our new solution methods make tractable. Figure 6 shows the total welfare change of moving to MP autarky in our calibrated models in pink while the blue bars decompose the total effect into the three channels of equation (9). Countries are ordered by the size of the total welfare effect in the calibration with positive complementarities. For each country, the left bars show results from the calibration with negative complementarities while the right bars show results from the calibration with positive complementarities.

Figure 6 shows that the vast majority of countries suffers negative welfare consequences from removing multinational production. Effects range from reductions in welfare of 20 log points for productive and open economies like the Netherlands, to negligible changes for some less productive and less open economies like Turkey. In general, the effect is less negative with negative complementarities compared to positive complementarities. For some less productive countries, such as Bulgaria, the opposite is true: with positive complementarities, there are welfare *gains* from moving to MP autarky while there are welfare losses under negative complementarities.

The countries on the left in Figure 6 are the main benefactors of multinational production: these are small, productive, and open economies in which many multinationals are headquartered. In these countries, like all countries, the openness channel decreases real consumption in the move to MP autarky. Moreover, their most productive firms lose the marginal cost advantages from multinational production, lowering average productivity substantially, especially with positive complementarities. In fact, some of these countries actually lose varieties, indicating that the low marginal costs attainable through foreign production were central in enabling firms with low productivity draws to make a positive profit; once these marginal cost advantages are no longer available, the number of surviving profitable domestic varieties shrinks.

By contrast, the countries on the right are small, low-wage locations with few domestic firms productive enough to engage in multinational production. Instead, they serve primarily as the production sites of productive firms headquartered elsewhere. The openness channel implies real consumption declines as the world production possibility shifts in and prices increase. However, these losses are offset by a positive variety effect: the labor released by departing multinationals crowds in the creation of new domestic varieties. The variety effect is stronger with positive complementarities, since multinational firms lose more of their marginal cost advantage when constrained to domestic-only production, shrinking and releasing more labor in the process. The average productivity effect is small for these countries, since the fact that so few domestic firms engage in multinational production implies that the term in square brackets is close to 1 as $s_{iii}^{t} \approx 1$ for all t. In turn, because relative firm sizes change little, the productivity cutoff remains (almost) unchanged and released labor predominantly increases entry.[17]

Overall, our analysis shows that calibrating the model with different types of complementarity *to the same data* yields substantially different welfare effects of MP autarky, with sign reversals in some cases.

Fixed costs are also essential: without fixed costs of establishing production locations, $\hat{\underline{z}}_i = 1$, and all firms produce in all countries. In this case, since all firms are affected symmetrically,

[17]If no domestic firm engages in multinational production, the distribution of domestic firms can be described by a representative firm as in the closed economy of Melitz (2003). In this case, as a consequence of CES demand together with Pareto productivity, changes in MP costs cause no reallocations among firms, so that the productivity of the representative firm does not change.

the term in square brackets collapses to a common share s_{iii} that captures labor reallocation between variable production and firm entry. In calibrations without fixed costs, welfare losses from MP autarky are substantially larger for nearly all countries, with either level of complementarity (see Figure 12 in the Online Appendix). Intuitively, with fixed costs, the retreat of multinational activity frees up labor that can be reallocated to the creation of new domestic varieties, partially offsetting the welfare loss. Thus, models that abstract from fixed costs tend to overstate the welfare losses from MP autarky.

In sum, our results suggest that modeling complementarities in location decisions—together with fixed costs—is crucial for quantitatively evaluating counterfactuals involving multinational production.

6. Conclusion

We introduce a method to solve combinatorial discrete choice problems with either negative or positive complementarities, and across heterogeneous agents. We apply it to a quantitative model of multinational production in which heterogeneous firms choose sets of foreign production locations. Our approach allows us to solve and calibrate the model in general equilibrium with many locations. The calibrated model shows that complementarities and fixed costs—central to the firms' combinatorial problem—are also critical for evaluating the gains from multinational production. Gains are larger with positive than with negative complementarities, and smaller when fixed costs are included. Models that abstract from these features may misstate the effects of disruptions that affect the cost of multinational production.

Beyond multinational production, our methods apply to a broad range of problems with interdependent discrete choices, for example in the context of discrete infrastructure investments, supply chain formation, or the location decisions of multi-establishment firms. By making such problems computationally tractable, our approach expands the frontier of applied general equilibrium analysis and enables counterfactuals that were previously deemed infeasible.

References

Adão, Rodrigo, Costas Arkolakis, and Sharat Ganapati (2020). *Aggregate Implications of Firm Heterogeneity: A Nonparametric Analysis of Monopolistic Competition Trade Models*. Tech. rep. National Bureau of Economic Research.

Alfaro-Urena, Alonso et al. (2023). *Firm Export Dynamics in Interdependent Markets*. Tech. rep. National Bureau of Economic Research.

Alviarez, Vanessa (2019). "Multinational Production and Comparative Advantage". In: *Journal of International Economics* 119, pp. 1–54.

Alviarez, Vanessa, Javier Cravino, and Natalia Ramondo (2023). "Firm-Embedded Productivity and Cross-Country Income Differences". In: *Journal of Political Economy* 131.9, pp. 2289–2327.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings (2019). "International Shocks, Variable Markups, and Domestic Prices". In: *The Review of Economic Studies* 86.6, pp. 2356–2402.

Antras, Pol, Teresa C Fort, and Felix Tintelnot (2017). "The Margins of Global Sourcing: Theory and Evidence from US Firms". In: *American Economic Review* 107.9, pp. 2514–64.

Antràs, Pol et al. (May 2024a). "Export-Platform FDI: Cannibalization or Complementarity?" In: *AEA Papers and Proceedings* 114.

— (Apr. 2024b). "Exporting, Global Sourcing, and Multinational Activity: Theory and Evidence from the United States". In: *The Review of Economics and Statistics*.

Arkolakis, Costas (2010). "Market Penetration Costs and the New Consumers Margin in International Trade". In: *Journal of Political Economy* 118.6, pp. 1151–1199.

— (2016). "A Unified Theory of Firm Selection and Growth". In: *The Quarterly Journal of Economics* 131.1, pp. 89–155.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, et al. (2019). "The Elusive Pro-Competitive Effects of Trade". In: *The Review of Economic Studies* 86.1, pp. 46–80.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare (2012). "New Trade Models, Same Old Gains?" In: *American Economic Review* 102.1, pp. 94–130.

Arkolakis, Costas, Fabian Eckert, and Rowan Shi (2023). *Combinatorial Discrete Choice: A Quantitative Model of Multinational Location Decisions*. Tech. rep. National Bureau of Economic Research.

Arkolakis, Costas, Natalia Ramondo, et al. (2018). "Innovation and Production in the Global Economy". In: *American Economic Review* 108.8, pp. 2128–73.

Behrens, Kristian et al. (2020). "Quantifying the Gap Between Equilibrium and Optimum under Monopolistic Competition". In: *The Quarterly Journal of Economics* 135.4, pp. 2299–2360.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile Prices in Market Equilibrium". In: *Econometrica*, pp. 841–890.

Boros, Endre and Peter L. Hammer (2002). "Pseudo-Boolean optimization". In: *Discrete Applied Mathematics* 123.1.

Broda, Christian and David E Weinstein (2006). "Globalization and the Gains from Variety". In: *The Quarterly Journal of Economics* 121.2, pp. 541–585.

Castro-Vincenzi, Juanma et al. (2024). "Differences in Product Availability Across Countries: Determinants and Welfare Consequences".

Chaney, Thomas (2008). "Distorted Gravity: the Intensive and Extensive Margins of International Trade". In: *American Economic Review* 98.4, pp. 1707–1721.

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

Costinot, Arnaud (2009). “An Elementary Theory of Comparative Advantage”. In: *Econometrica* 77.4.

De Loecker, Jan et al. (2016). “Prices, Markups, and Trade Reform”. In: *Econometrica* 84.2, pp. 445–510.

Dyrda, Sebastian, Guangbin Hong, and Joseph B. Steinberg (2024). “A Macroeconomic Perspective on Taxing Multinational Enterprises”. In: *Journal of International Economics* 152.

Eaton, Jonathan and Samuel Kortum (2002). “Technology, Geography, and Trade”. In: *Econometrica* 70.5, pp. 1741–1779.

Fajgelbaum, Pablo D et al. (2019). “State Taxes and Spatial Misallocation”. In: *The Review of Economic Studies* 86.1, pp. 333–376.

Fan, Ying and Chenyu Yang (2020). “Competition, Product Proliferation, and Welfare: A Study of the US Smartphone Market”. In: *American Economic Journal: Microeconomics* 12.2, pp. 99–134.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer (2015). “The Next Generation of the Penn World Table”. In: *American Economic Review* 105.10, pp. 3150–3182.

Gentzkow, Matthew (June 2007). “Valuing New Goods in a Model with Complementarity: Online Newspapers”. In: *American Economic Review* 97.3.

Goldberg, Pinelopi Koujianou (1995). “Product Differentiation and Oligopoly in International Markets: The Case of the US Automobile Industry”. In: *Econometrica*, pp. 891–951.

Guillouët, Louise et al. (Aug. 2024). *Language Barriers in Multinationals and Knowledge Transfers*. Tech. rep.

Head, Keith and Thierry Mayer (2014). “Gravity Equations: Workhorse, Toolkit, and Cookbook”. In: *Handbook of International Economics*. Vol. 4. Elsevier, pp. 131–195.

— (2019). “Brands in Motion: How Frictions Shape Multinational Production”. In: *American Economic Review* 109.9, pp. 3073–3124.

Head, Keith, Thierry Mayer, et al. (2024). “Industrial Policies for Multi-Stage Production: The Battle for Battery-Powered Vehicles”.

Hendel, Igal (1999). “Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns”. In: *The Review of Economic Studies* 66.2, pp. 423–446.

Hjort, Jonas, Hannes Malmberg, and Todd Schoellman (2022). *The Missing Middle Managers: Labor Costs, Firm Structure, and Development*. Tech. rep. National Bureau of Economic Research.

Jia, Panle (2008). “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry”. In: *Econometrica* 76.6, pp. 1263–1316.

Jiang, Xian (2023). *Information and Communication Technology and Firm Geographic Expansion*. Tech. rep.

Kleene, Stephen C. (Dec. 1936). "General Recursive Functions of Natural Numbers". In: *Mathematische Annalen* 112, pp. 727–742.

— (1938). "On Notation for Ordinal Numbers". In: *Journal of Symbolic Logic* 3.4, pp. 150–155.

Klenow, Peter J and Jonathan L Willis (2016). "Real Rigidities and Nominal Price Changes". In: *Economica* 83.331, pp. 443–472.

Klimeš, Jiří (1981). "Fixed Edge Theorems for Complete Lattices". In: *Archivum Mathematicum* 17.4, pp. 227–234.

Kulesza, Adrian (2024). *Combinatorial Discrete Choice with Deep Reinforcement Learning*.

Lind, Nelson and Natalia Ramondo (Feb. 2023). "Trade with Correlation". In: *American Economic Review* 113.2, pp. 317–53.

Liu, Jin (2023). *Multinational Production and Innovation in Tandem*. Tech. rep. Working Paper. New York University.

McFadden, Daniel (1974). "Conditional Logit Analysis of Qualitative Choice Behavior". In: *Frontiers in Econometrics*.

Melitz, Marc J. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity". In: *Econometrica* 71.6, pp. 1695–1725.

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

Milgrom, Paul and Chris Shannon (1994). "Monotone Comparative Statics". In: *Econometrica: Journal of the Econometric Society*, pp. 157–180.

Muendler, Marc-Andreas and Sascha O. Becker (Dec. 2010). "Margins of Multinational Labor Substitution". In: *American Economic Review* 100.5, pp. 1999–2030.

Pollak, Robert A (1971). "Additive Utility Functions and Linear Engel Curves". In: *The Review of Economic Studies* 38.4, pp. 401–414.

Ramondo, Natalia (2014). "A Quantitative Approach to Multinational Production". In: *Journal of International Economics* 93.1, pp. 108–122.

Ramondo, Natalia and Andrés Rodríguez-Clare (2013). "Trade, Multinational Production, and the Gains from Openness". In: *Journal of Political Economy* 121.2, pp. 273–322.

Ramondo, Natalia, Andrés Rodríguez-Clare, and Felix Tintelnot (2015). "Multinational Production: Data and Stylized Facts". In: *American Economic Review* 105.5, pp. 530–536.

Sabal, Alejandro (2025). *Product Entry in the Global Automobile Industry*. Tech. rep.

Silva, JMC Santos and Silvana Tenreyro (2006). "The Log of Gravity". In: *The Review of Economics and Statistics* 88.4, pp. 641–658.

Simonovska, Ina (2015). "Income Differences and Prices of Tradables: Insights from an Online Retailer". In: *The Review of Economic Studies* 82.4, pp. 1612–1656.

Sotelo, Sebastian (2019). "Practical Aspects of Implementing the Multinomial PML Estimator". In: *Ann Arbor: University of Michigan, mimeo*.

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

Tarski, Alfred (1955). "A Lattice-Theoretical Fixpoint Theorem and its Applications". In: *Pacific Journal of Mathematics* 5.2, pp. 285–309.

Tintelnot, Felix (2017). "Global Production with Export Platforms". In: *The Quarterly Journal of Economics* 132.1, pp. 157–209.

Train, Kenneth E., Daniel McFadden, and Moshe Ben-Akiva (1987). "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices". In: *The RAND Journal of Economics* 18.1.

Xiang, Wei (2022). "Clean Growth and Environmental Policies in the Global Economy". Yale University mimeo.

Zheng, Fanyin (2016). "Spatial Competition and Preemptive Entry in the Discount Retail Industry". In: *Columbia Business School Research Paper* 16-37.

A. Counterfactual welfare

The trilateral sales in the model are given by

$$X_{i\ell n} = \tilde{\zeta} M_i X_n \left(\frac{P_n}{\mu}\right)^{\sigma-1} \sum_{\mathcal{Z}_i^t \in \mathcal{T}_i} \mathbb{1}_{i\ell}^t \xi_{i\ell n}^{-\theta} \left(\Theta_{in}^t\right)^{\frac{\sigma-1}{\theta}-1} \underline{z}_i^{\zeta} \left[\left(z_i^{t+1}\right)^{\sigma-1-\zeta} - \left(z_i^t\right)^{\sigma-1-\zeta}\right]$$

where $\tilde{\zeta}$ is a constant of integration and, for brevity, we denote $\Theta_{in}^t \equiv \sum_{\ell \in \mathcal{L}_i^t} \xi_{i\ell n}^{-\theta}$. Then, the counterfactual change is

$$\hat{X}_{i\ell n} = \hat{M}_i \hat{w}_n \hat{P}_n^{\sigma-1} \frac{\sum_{\mathcal{Z}_i^{t\prime} \in \mathcal{T}_i^{\prime}} \mathbb{1}_{i\ell}^{t\prime} \xi_{i\ell n}^{\prime-\theta} \left(\Theta_{in}^{t\prime}\right)^{\frac{\sigma-1}{\theta}-1} \left[\left(z_i^{t+1\prime}\right)^{\sigma-1-\zeta} - \left(z_i^{t\prime}\right)^{\sigma-1-\zeta}\right]}{\sum_{\mathcal{Z}_i^t \in \mathcal{T}_i} \mathbb{1}_{i\ell}^t \xi_{i\ell n}^{-\theta} \left(\Theta_{in}^t\right)^{\frac{\sigma-1}{\theta}-1} \left[\left(z_i^{t+1}\right)^{\sigma-1-\zeta} - \left(z_i^t\right)^{\sigma-1-\zeta}\right]}.$$

Computing the counterfactual change in X_{iii} derives the welfare formula. To do so, we adopt the following assumption for the remainder of the section.

Assumption. *Model fundamentals are such that, in the baseline equilibrium, it is the case that* $\mathbb{1}_{ii}(z) = 1$ *for all* $z \geq \tilde{z}_i$*, that is, all active firms establish production in their country of origin* i*.*

This assumption guarantees that no firms produce in foreign locations but not domestically. With either MP or trade autarky, the change simplifies to

$$\begin{aligned}\hat{X}_{iii} &= \hat{M}_i \hat{w}_i \hat{P}_i^{\sigma-1} \hat{\xi}_{iii}^{-\theta} \frac{\left(\xi_{iii}^{\prime-\theta}\right)^{\frac{\sigma-1}{\theta}-1} \left(\tilde{z}_i^{\prime}\right)^{\sigma-1-\zeta}}{\sum_{\mathcal{Z}_i^t \in \mathcal{T}_i} \mathbb{1}_{ii}^t \left(\Theta_{ii}^t\right)^{\frac{\sigma-1}{\theta}-1} \left[\left(z_i^t\right)^{\sigma-1-\zeta} - \left(z_i^{t+1}\right)^{\sigma-1-\zeta}\right]} \\ &= \hat{M}_i \hat{w}_i \hat{P}_i^{\sigma-1} \hat{\xi}_{iii}^{1-\sigma} \hat{\tilde{z}}_i^{\sigma-1-\zeta} \sum_{\mathcal{Z}_i^t \in \mathcal{T}_i} \lambda_{iii}^k \left(s_{iii}^t\right)^{\frac{\sigma-1}{\theta}-1},\end{aligned}$$

where we use the fact that $\hat{\xi}_{i\ell n}^{-\theta} = 0$ for all $i \neq \ell$ and define

$$s_{i\ell i}^t \equiv \frac{\xi_{i\ell i}^{-\theta}}{\sum_{\mathcal{L}_i^t} \xi_{i\ell i}^{-\theta}} \qquad , \qquad \lambda_{iii}^k \equiv \frac{\mathbb{1}_{ii}^t s_{iii}^t \left(\Theta_{ii}^t\right)^{\frac{\sigma-1}{\theta}} \left[\left(z_i^t\right)^{\sigma-1-\zeta} - \left(z_i^{t+1}\right)^{\sigma-1-\zeta}\right]}{\sum_{\mathcal{Z}_i^t \in \mathcal{T}_i} \mathbb{1}_{ii}^t s_{iii}^t \left(\Theta_{ii}^t\right)^{\frac{\sigma-1}{\theta}} \left[\left(z_i^t\right)^{\sigma-1-\zeta} - \left(z_i^{t+1}\right)^{\sigma-1-\zeta}\right]}.$$

The first is the share of production, sold in i and produced by interval $\mathcal{Z}_i^t$, that takes place in location ℓ compared to other locations. The second is the share of production X_{iii} that is accounted for by interval $\mathcal{Z}_i^t$. Using the fact that $\hat{\xi}_{iii}^{1-\sigma} = \hat{w}_i^{1-\sigma}$, then rearranging, we arrive at

$$\frac{\hat{w}_i}{\hat{P}_i} = \hat{\pi}_{iii}^{\frac{1}{1-\sigma}} \hat{M}_i^{\frac{1}{\sigma-1}} \hat{\bar{z}}_i^{1-\frac{\zeta}{\sigma-1}} \left[\sum_{\mathcal{Z}_i^t \in \mathcal{T}_i} \lambda_{iii}^t \left(s_{iii}^t \right)^{\frac{\sigma-1}{\theta}-1} \right]^{\frac{1}{\sigma-1}} .$$

B. Input Microfoundation for CES Marginal Cost

In this section, we lay out a microfounded motivation for the CES marginal cost function employed in the main body of the paper that follows Antras, Fort, and Tintelnot (2017).

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

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

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

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

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

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

C. Data and Calibration

In this section, we discuss our data construction in more detail and provide additional information about the calibration.

C.1. Data

Trade, Foreign Affiliate Sales, and Foreign Affiliate Counts We use the data set compiled by Alviarez (2019) for our bilateral flow data. The data set combines information from four major databases: OECD International Direct Investment Statistics and the Statistics on Measuring Globalization; Eurostat Foreign Affiliate Statistics database; Bureau of Economic Analysis (BEA) public data; and Bureau van Dijk's Orbis dataset. All values in the data set are averages from 2003 to 2012; accordingly, in all other data sets we use in the calibration, we also take average values over the same period.

The data contains information for the following set of 32 countries for nine sectors. While we keep all the countries, we collapse all data across manufacturing industries to obtain manufacturing sector totals. For all combinations of these countries and sectors, the data contains the value of trade flows. We construct home absorption by summing a country's total export sales and subtracting them from the total sales of the sector in the country to obtain sales to the domestic market. Using these home sales, we can then construct home trade shares.

In addition, for each such origin-destination-sector triplet, the data contain the total sales of foreign affiliates, e.g. the total sales of Canadian companies located in Germany; note that there is no information on the destination countries of foreign affiliates in a given country, i.e. the data set does not report how much Canadian companies in Germany are selling to Greece. We construct sales of a country's domestic firms at home to destinations anywhere in the world by taking the total sales of all firms in the country and subtracting the total sales by foreign affiliates of other countries done in the country. Using these home sales, we can construct the complete matrix of (inward and outward) MP.

Lastly, the data contain information on the total foreign affiliates for each country-sector pair. The data do not contain information on the total domestic enterprises. We bring in data on the average number of total enterprises in each country between 2003 and 2012 (see below). We then subtract the total number of foreign affiliates operating in a country from the total number of enterprises and interpret the difference as the number of domestic enterprises or "headquarters" in our model. We can then once again compute the entire matrix of (inward and outward) MP enterprise shares for all country combinations.

CEPII Data We use the standard set of gravity variables from the CEPII database (see Conte, Cotterlaz, Mayer, et al. 2023) with minor modifications. As our distance measure, we use the simple distance between countries' most populated cities, measured in kilometers. To generate our colonial dummy variable, we combine the "colonial sibling" dummy, which indicates if two countries had a past common colonizer, and the "colonial dependence" dummy, which indicates if one country ever colonized the other from the CEPII data into one "colonial relationship" dummy. We also use the two dummy variables that indicate for every country pair whether it shares a border or an official language.

TRAINS Tariff Data We use information on tariffs from the TRAINS dataset for 2003–2012. This database reports the MFN (most favored nation) tariff rates for over 5,000 HS6 goods categories and country pair combinations. The data also contain information on preferential trade agreements and their postulated rates. We drop observations for which trade is subject to non-ad valorem (specific or nonlinear/compound) tariffs. For these tariffs, TRAINS reports ad-valorem equivalents. However, computation of these equivalents requires data on quantities, which are often noisy and could also endogenously respond to changes in tariffs. Since most MFN tariffs are ad valorem, the impact of dropping these observations for our sample size is small. For each country and HS pair, we take the minimum of the MFN and preferred rate, and then we take an unweighted average of the resulting tariffs across all HS codes in the NAICS-33 (manufacturing) sector and all years for which the tariffs are not missing.

We do several robustness tests for our tariff data: First, we use MFN tariffs since, at times, even if a preferred agreement is in place, firms trade using MFN rates since trading subject to preferred rates may require extra efforts for firms, e.g. additional forms to fill out. Second, we experiment with the weighting of HS goods and construct a weighted tariff for manufacturing that is weighted by the goods share in world trade in that year. Third, we use the information on applied rates, which is available for goods traded in positive quantities only. We assign the applied tariff to a good-country pair if available and otherwise assign the minimum of MFN and preferred rate. None of these alternative measures alters our quantitative results significantly.

OECD Enterprise Data We obtain information on the number of enterprises active in each country from the OECD Structural Statistics of Industry and Services for OECD and non-OECD countries. We also extract additional data from the National Statistical Agencies for Ukraine, Mexico, and Russia, for which the OECD data lacks information.

OECD Survival Statistics We also use the OECD Structural and Demographic Business Statistics database to obtain information on the 1-, 2-, 3-, 4-, and 5-year survival rates of manufacturing enterprises in many countries and years. The data is missing for some year and country combinations. We impute the missing survival rates by running a regression of log survival rates on the log of GDP, total employment, total population, and year and survival rate horizon fixed effects. We then use the estimated coefficients to predict the missing survival rates. In our baseline calibration, we use the 1-year survival rates since they correspond most closely to the idea of businesses that pay an entry cost to learn their productivity but then never end up producing positive quantities.

Penn World Tables We use the PWT 10.01 version of the Penn World Tables (see Feenstra, Inklaar, and Timmer 2015). We extract the total "expenditure side real GDP in chained PPPs in millions of 2017 dollars between 2003 and 2012. Likewise, we extract the total employment and total population of each country in each year.

C.2. Calibration

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

Bilateral Costs of Trade, MP, and Affiliates We estimate the following empirical gravity equation for each bilateral flow indexed by $x \in \{\tau_{\ell n}, \gamma_{i\ell}, \nu_{i\ell}\}$: trade flows, inward MP sales, and inward affiliate stocks.

$$y_{ij}^x = \exp\left(\alpha^x + \sum_{v \in \{d, \text{COL}, \text{COM}, \text{BOR}\}} \beta_v^x v_{ij} + \delta_x' X_{ij} + \varkappa_i + \zeta_j\right) + \epsilon_{ij}^x \tag{10}$$

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

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

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

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

where v indexes the same set of gravity variables as in the regression (10). The $\{\overline{\tau}_n, \overline{\gamma}_\ell, \overline{\nu}_\ell\}$ components represent the costs of doing an activity across borders versus within borders. In our calibration procedure, we estimate the destination-specific components $\{\overline{\tau}_n, \overline{\gamma}_\ell, \overline{\nu}_\ell\}$ by targeting the own-shares of each activity, and the gravity-variable-specific elasticities by targeting the estimated coefficients on the corresponding gravity variables in Table 3a using unweighted tariffs. Table 7a reports the estimated elasticities.

Figure 7b shows histograms of our calibrated trade costs, MP costs, and the bilateral component of fixed costs; we exclude the diagonal entries of all cost matrices from the histogram since they are normalized to 1. Our estimated trade costs are substantial, similar to prior estimates from studies featuring trade and multinational production (e.g. Ramondo and Rodríguez-Clare 2013).

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

In the Online Appendix, we regress our estimates of the MP costs and the bilateral component of the fixed costs on our set of gravity variables to understand their determinants. For the bilateral component of fixed costs, we find a large positive coefficient on language, consistent with evidence of language barriers in foreign MNE activity by Guillouët et al. (2024). Our cost estimates are also consistent with the finding in Alviarez, Cravino, and Ramondo (2023) that within-destination market firm market shares for manufacturing decline relatively little with distance. In particular, MP costs, which are the main determinant of within-destination market shares in our model, react little to distance in our calibrations.

Model Fit Figure 8 presents our calibrated model's performance on an important untargeted moment: the US sales premium of multinational firms based in the US. We calculate the average sales among groups of firms with presence in different minimum numbers of foreign locations, normalized by the average sales of non-MNE firms based in the US. Figure 8 shows both these MNE sales premia computed in our calibrated model and the empirical premia documented by Antràs et al. (2024b). The model premia closely mirror the empirical premia, reflecting that more productive firms broadly have both more foreign affiliate locations and higher sales. For example, MNEs (any firm conducting production in at least two countries) are more than 40 times larger than non-MNEs in the US in both the model and the (untargeted) data.

	Unweighted Tariffs			Weighted Tariffs		
	Trade (1)	MP (2)	Affiliates (3)	Trade (4)	MP (5)	Affiliates (6)
Log Distance	-0.687***	-0.289**	-0.681***	-0.689***	-0.284**	-0.681***
	(0.0541)	(0.106)	(0.0847)	(0.0540)	(0.107)	(0.0851)
Colony	0.0776	-0.00471	0.232	0.0758	0.000476	0.260
	(0.125)	(0.131)	(0.144)	(0.124)	(0.131)	(0.146)
Contiguity	0.448***	0.424**	0.448***	0.440***	0.412*	0.436***
	(0.0697)	(0.164)	(0.0979)	(0.0697)	(0.163)	(0.0983)
Language	0.152	0.468**	0.561***	0.160	0.476**	0.577***
	(0.102)	(0.145)	(0.155)	(0.101)	(0.146)	(0.161)
Observations	992	707	710	992	707	710

(A) Estimation with PPML

	Unweighted Tariffs			Weighted Tariffs		
	Trade (1)	MP (2)	Affiliates (3)	Trade (4)	MP (5)	Affiliates (6)
Log Distance	-1.106***	-0.822***	-0.800***	-1.104***	-0.814***	-0.796***
	(0.0510)	(0.122)	(0.0740)	(0.0509)	(0.122)	(0.0741)
Colony	0.763***	0.940***	0.659***	0.761***	0.954***	0.669***
	(0.108)	(0.244)	(0.149)	(0.108)	(0.244)	(0.149)
Contiguity	0.325***	0.734***	0.410***	0.325***	0.732***	0.408***
	(0.0913)	(0.195)	(0.118)	(0.0912)	(0.195)	(0.118)
Language	0.00442	0.559*	0.183	0.0000334	0.570*	0.190
	(0.134)	(0.284)	(0.173)	(0.133)	(0.284)	(0.173)
Observations	992	707	710	992	707	710

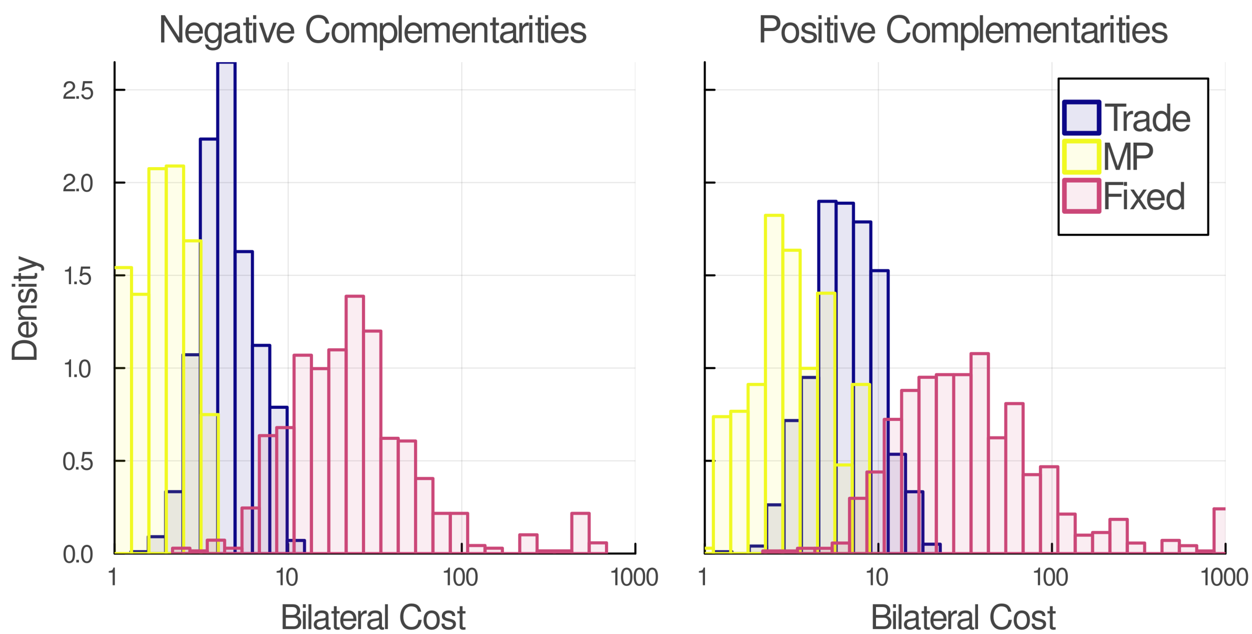
(B) Estimation with OLS

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

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

	Negative Complementarities			Positive Complementarities		
	Trade	MP	Affiliates	Trade	MP	Affiliates
Language	0.047	0.066	0.080	0.064	0.106	0.197
Contiguity	0.143	0.060	0.053	0.162	0.134	-0.010
Colony	0.025	-0.028	0.187	0.025	-0.039	0.207
Log Distance	0.215	0.001	0.346	0.270	0.016	0.429

(A) Estimated Cost Elasticities of the Gravity Variables



(B) Distribution of Calibrated Bilateral Costs

FIGURE 7: TRADE COSTS, MP COSTS, AND THE BILATERAL COMPONENT OF FIXED COSTS

This figure summarizes the estimated bilateral costs. Figure 7b shows a histogram of the three bilateral cost matrices in the model: trade costs, MP costs, and the bilateral component of fixed costs. We omit the own-country costs which are normalized to 1 for all three types of costs. For MP and the bilateral component of fixed costs, we also omit country pairs where MP is zero, since we set the MP costs to be infinity in those cases. Table 7a presents the calibrated elasticities of all gravity variables in each of the bilateral costs in the model as specified in equation (11).

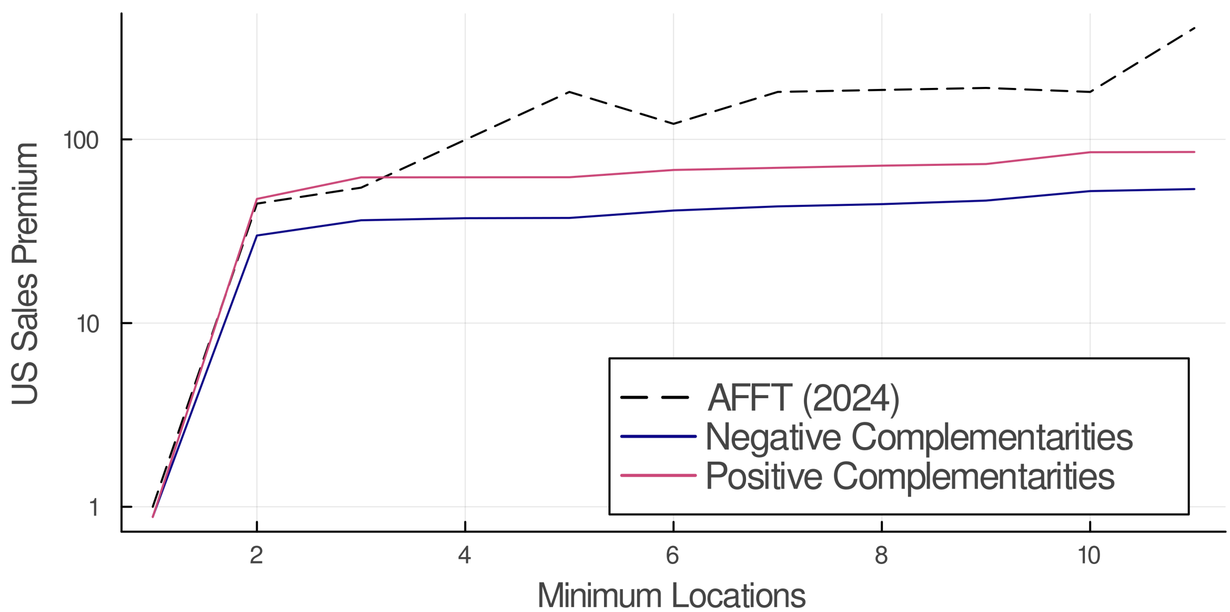


FIGURE 8: MULTINATIONAL SALES PREMIA AND NUMBER OF FOREIGN AFFILIATES IN THE DATA AND THE BASELINE CALIBRATION

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