Combinatorial Discrete Choice: A Quantitative Model of Multinational Location Decisions*

Costas Arkolakis† Fabian Eckert‡ Rowan Shi§

First Version: January 2017
This Version: November 2023

Abstract

We introduce a general quantifiable framework to study the location decisions of multinational firms. In the model, firms choose in which locations to pay the fixed costs of setting up production, taking into account potential complementarities among production locations. The firm's location choice problem is combinatorial because the marginal value of an individual production location depends on its complete set of production sites. We develop a computational method to solve such problems and aggregate optimal decisions across heterogeneous firms. We use our calibrated model to study Brexit and the recent sanctions war with Russia. In both counterfactuals, changes in the location decisions of multinationals are driving real wage responses.

*This paper combines "Combinatorial Discrete Choice" by Arkolakis and Eckert, and the technical part of the first chapter of Shi's dissertation titled "Solving combinatorial discrete choice problems in heterogeneous agent models: theory and an application to corporate tax harmonization in the European Union." We thank Jake DiCicco, Tra Nguyen, Andrew Salmon, Benjamin Tong, Yunus Tuncbilek, and Nardeen Abdulkareem for outstanding research assistance. We also thank Chrysanthos Gounaris for early conversations on the topic as well as David Atkin, Arnaud Costinot, Dave Donaldson, Penny Goldberg, Navin Kartik, Samuel Kortum, Eduardo Morales, Nitya Pandalai-Nayar, and Steve Redding for comments and suggestions. Shi gratefully acknowledges support from the Princeton International Economics Section. Code implementing the techniques introduced in this paper can be found at https://github.com/rowanxshi/CDCP.jl.

†Yale University, Department of Economics, and NBER, Email: costas.arkolakis@yale.edu
‡University of California, San Diego, Department of Economics, Email: fpe@ucsd.edu
§Toronto Metropolitan University, Department of Economics, Email: rowanxshi@torontomu.ca

1 Introduction

In 2016, multinational enterprises (MNEs) accounted for about a quarter of global value-added and half of the world's export flows in manufacturing. At the same time, only a tiny fraction of existing manufacturing firms are MNEs, typically less than one percent.[1] In light of these facts, fixed costs of setting up foreign operations are a central building block in models of MNE activity (Yeaple (2013); Helpman et al. (2004); Antràs and Yeaple (2014)). In the context of a quantitative model with realistic geography, the resulting scale economies imply that the value of any individual plant depends on the firm's entire set of plant locations; locations are non-fungible since each is unique in its geography. As a result, an MNE's profit maximization problem involves evaluating all potential combinations of production locations, which is computationally infeasible even with a moderate number of locations. We refer to such multiple discrete choice problems as combinatorial discrete choice problems, or CDCPs.

This paper provides a quantitative framework to study the location decisions of multinational firms in an environment where setting up foreign production sites incurs fixed costs and there are complementarities among sites. At the heart of our paper are new iterative methods to solve the resulting combinatorial discrete choice problem confronted by firms and aggregate their optimal location decisions into aggregate outcomes. We use our solution methods to calibrate the model to study the exit decision of Britain from the European Union (EU). Changes in the production location decisions of multinationals imply sizable losses for Britain, some losses for EU countries, and gains for non-EU members. We also study sanctions on Russia in the aftermath of the invasion of Ukraine and find significant changes to trade flows and multinational production volumes, as well as changes in the location decisions of MNEs that affect nearby countries the most.

In our theoretical framework, firms headquartered in one country pay destination-specific fixed costs to set up production in other countries. Foreign production locations may be less productive than domestic ones but allow firms to save on labor and trade costs to foreign destinations. A firm's final good consists of a unit continuum of intermediate varieties that all its sites can produce. The firm finds the cheapest supplier among its production locations for each destination and intermediate input variety. As a result, a firm's production sites cannibalize one another's sales, giving rise to a negative "supply-side" complementarity among locations. On the demand side, a general demand function for the final product of each firm gives rise to a positive complementarity among production locations. In particular, firms with more

[1] These figures are constructed using the OECD AMNE database for a set of 36 OECD and 23 non-OECD countries. The data come from Alviarez (2019) and authors' own calculations (see Section 5). MP activity is even more concentrated than exporting activity. Bernard et al. (2007) report that only 18 percent of all US manufacturing firms had positive exports in 2002; Boehm et al. (2023) show that this number had increased to 23 percent by 2012; Eaton et al. (2011) find similar numbers for the French economy.

production sites have lower marginal costs and higher sales; each additional location at such firms is more valuable because the associated marginal cost savings apply to a larger sales volume.

Together, the fixed costs of setting up a production location and the complementarities among these locations make MNEs' location decisions combinatorial, because the value of any particular location depends on the complete set of production sites. Such problems quickly become intractable as the space of possible combinations of locations grows exponentially in their total number. We present an iterative procedure to solve such problems even with many locations. Our method requires the firm's profit function to exhibit a "single crossing differences in choices" property (SCD-C), which restricts the complementaries among locations to be either always weakly positive or always weakly negative.[2] Our method iteratively eliminates many non-optimal plant location sets without explicitly computing their associated profits by exploiting the monotonicity implied by the SCD-C condition.

Another challenge of solving CDCPs in quantitative models is aggregating optimal behavior across heterogeneous firms, whose optimal production location sets may vary. We introduce a set-valued "policy function" that maps each firm type into an optimal set of production locations and show that it can exhibit jumps, non-monotonicities, and partially overlapping production location sets. Computing general equilibrium objects then requires integrating the policy function over the distribution of firm types. To recover the policy function, we provide an iterative technique that applies as long as the firm's profit function additionally satisfies a single crossing differences in type (SCD-T) condition. The SCD-T condition restricts how profits vary with firm type and guarantees that similar firms choose the same production sites so that the policy function only changes value at discrete "cutoffs" in the type space. Our aggregation method identifies these cutoffs and the optimal production location sets shared by the types between them.

Our solution and aggregation methods apply to general combinatorial discrete choice problems outside the particular context of multinational production as long as the objective function satisfies the relevant single-crossing conditions. To apply the methods to our model, we provide sufficient conditions under which the firm's profit function satisfies the SCD-C and SCD-T conditions. For SCD-C, we show that the strength of negative complementarities depends on the elasticity of substitution among a firm's production locations. The strength of the positive complementarities depends on the elasticities of demand and pass-through, which govern how marginal cost savings through an additional production location affect a firm's overall profits. Whether the firm's production location problem exhibits positive or negative complementarities then depends on the relative size of these elasticities. We also extend this intuition to establish

[2]The “Single Crossing Differences” property first appeared in Milgrom (2004). While well-known in the microeconomics literature, to our knowledge it has not been discussed in the context of solving combinatorial discrete choice problems.

SCD-C and SCD-T in frameworks with more general correlated or multidimensional technology shocks, as in Ramondo and Rodríguez-Clare (2013), Lind and Ramondo (2023), and Xiang (2022), and more general variable elasticity demand functions, as in Arkolakis et al. (2019) and Matsuyama and Ushchev (2017).

For our quantitative exercise, we choose a version of our model which features a Constant Elasticity of Substitution (CES) demand system. In this case, the firm's location choice problem always satisfies SCD-C, and the sign of the complementarities among locations depends on the relative size of two parameters: the price elasticity of final demand and the substitutability of production locations in the firm's production function. Most estimates of these parameters from the multinational production literature imply that the negative supply-side complementarities dominate the firm's location choice problem. We adopt central values of these parameters for our baseline calibration but also provide robustness exercises with values that imply starker negative or positive complementarities. Unlike standard quantitative trade models, fixed costs and discrete decisions imply that our model does not deliver closed-form aggregate gravity relationships. We calibrate trade costs, productivity costs of foreign production, and the fixed costs of foreign production to match the bilateral matrices of trade flows, multinational sales, and foreign affiliate counts across 32 countries using a rich data set produced by Alviarez (2019).

With the calibrated model, we illustrate the performance of our algorithm in a number of speed tests. Computation time increases polynomially in the number of countries with our algorithm but exponentially when evaluating all possible location strategies individually ("brute force"). The main alternative to our method of recovering the policy function is to discretize firm heterogeneity and then interpolate optimal location decision sets between grid points. We show that our solution method for the policy function yields machine precision solutions as fast as common discretization approaches require to find solutions with a high degree of discretization error. Overall, our method is drastically faster than existing approaches and can precisely solve previously infeasible problems with many locations and negative complementarities among production locations.

Finally, we use the model to simulate two recent major political events with economic consequences specifically relating to the operations of multinational firms: Brexit and the sanctions placed on Russia in the aftermath of the invasion of Ukraine. We conduct counterfactual exercises suggestive of the general equilibrium impacts of these events Using the estimated model, we progressively simulate increasing frictions between Great Britain and European Union member states. First, we increase trade costs by 10%; then we lower the relative efficiency of foreign production sites by 10%; and, finally, we increase the fixed costs of foreign production by 10%. When only trade costs increase, a standard proximity-concentration trade-off (Brainard (1997); Tintelnot (2017); Helpman et al. (2004)) increases the bilateral multinational activity between Great Britain and EU countries. However, prices increase in tandem because of the higher trade costs, resulting in small negative effects on real wages. Once we also lower the relative efficiency of foreign production sites, EU-based firms shut down production sites in Great Britain, resulting in substantially lower nominal wages and higher prices in not only Great Britain but also its major trading partners, such as the Netherlands and Belgium.

To simulate a counterfactual that resembles the sanction war with Russia, we lower the relative efficiency of foreign production sites and increase the fixed costs of foreign production between Russia and countries in the EU and the USA by 30% relative to their values in our baseline calibration. The sanctions lead to large reallocations of production sites and multinational activity. The effects of geography are salient, as firms from non-sanction countries with headquarters closer to Russia increase their production in Russia by more than firms headquartered in countries far away. Countries that impose sanctions and are close to Russia contract production and shut down production locations by much more than sanctioning countries that are further away.

Our paper contributes to a growing literature in international trade and industrial organization in which firms choose a discrete set of locations to produce in or source inputs from. The existing literature either abstracts from fixed costs and, thus, CDCPs altogether (see Ramondo (2014), Ramondo and Rodríguez-Clare (2013), Arkolakis et al. (2018), Fajgelbaum et al. (2019)), solves CDCPs by evaluating all possible location choices which is feasible only with a small number of locations (see Tintelnot (2017), Zheng (2016), Dyrda et al. (2023)), or restricts the analysis to CDCPs with the easier-to-solve case of positive complementarities (see Antras et al. (2017)).[3] Our methods open the way to solve, calibrate, and simulate large quantitative general equilibrium models of multinational production featuring heterogeneous firms, fixed costs, and positive or negative complementarities among production locations. Fast and exact aggregation allows estimating such models using aggregate moments, which requires solving for general equilibrium repeatedly.

We also contribute to a literature studying (combinatorial) discrete choice problems by introducing a unified way of solving single-agent CDCPs with either positive or negative complementarities. In most classical discrete choice problems, agents choose one item from a set of mutually exclusive alternatives (e.g., MCFADDEN (1974)). In some applications, agents select multiple items without complementarities among the individual items, but with a pre-specified total number of choices (e.g., Hendel (1999)). Fan and Yang (2020) propose a heuristic "greedy algorithm" to solve CDPCs without restrictions on the type of complementarities which works well to find *local* solutions, but, in contrast to our method, does not provide theoretical guarantees for a *global* optima. Jia (2008) is an exception: the paper provides a method to find the global solution to combinatorial discrete choice problems in which the objective exhibits

[3] A recent paper by Oberfield et al. (2023) solves combinatorial location problems using a continuous limit; their approach is complementary to our discrete locations approach. Other papers have taken the approach of estimating the parameters of the CDCP using moment inequalities and use them for counterfactuals that hold location choices fixed (see Morales et al. (2019), Holmes (2011)).

positive complementarities. Relative to Jia (2008), we show how to solve CDCPs with negative complementarities which naturally occur in models of multinational production and how to aggregate the solution to CDCPs across heterogeneous firms.

A distinct contribution is our method to recover the policy function that maps agent types to optimal location sets and is central to aggregating firm decisions into general equilibrium objects. Aggregation of choices in classical discrete choice models usually relies on a random utility shocks (see Guadagni and Little (1983) and Train (1986)) which require agents to choose only a single option among discrete alternatives. In the context of multiple-discrete-choice problems with positive complementarities, existing approaches have relied on discretizing the type heterogeneity, solving the CDCP for each grid point, and then interpolating between grid points (see Antras et al. (2017)). We show that such interpolation can lead to large errors in settings with negative complementarities since the relevant policy function exhibits no continuity or nesting. We provide a new method to solve for the exact policy function that takes roughly the same time as discretized approaches.

2 A Quantitative Model of Multinational Production

In this section, we introduce our model of multinational production. We characterize the firms' multinational location problem and the economic mechanisms that make it combinatorial. In the Online Appendix, we generalize this 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, locations of production by ℓ, and locations of final consumption by n. In each production location ℓ, there is a mass of agents H_ℓ, who each inelastically provide one unit of labor to firms at wage w_ℓ. Firms headquartered in location i each produce a single differentiated final product $\omega \in \Omega_i$, where Ω_i is the set of final goods produced in location i. Labor markets are perfectly competitive and output markets are monopolistically competitive. Figure 1 provides a schematic overview of our theoretical framework.

Demand System All consumers have identical preferences over the set of final goods $\Omega = \Omega_1 \cup \ldots \cup \Omega_{|L|}$. As in Arkolakis et al. (2019), for a given schedule of prices $\{p(\omega)\}_{\omega\in\Omega}$, consumers in market n demand the following quantity of good ω:

$$q_n\left(p\left(\omega\right)\right) = Q_n D\left(p\left(\omega\right)/P_n\right), \tag{1}$$

where Q_n and P_n are aggregate demand shifters specific to market n and $D(\cdot)$ is a positive, strictly decreasing, and differentiable function. As discussed in Arkolakis et al. (2019), equation

(1) can result from a variety of homothetic or non-homothetic utility functions.[4] Firms take the aggregate objects Q_n and P_n as given. The first part of our analysis does not require further restrictions on the aggregators P_n and Q_n.

Production Technology Consider a firm headquartered in location i that produces the final good ω and operates a set of production locations $\mathcal{L} \subseteq L$ which we take as given for now. The firm is associated with a vector $\mathbf{z}(\omega) = \{z_1(\omega), \ldots, z_{|L|}(\omega)\}$, which collects its productivity of producing good ω in any potential location ℓ. Firms with the same headquarter location and the same productivity vector $\mathbf{z}(\omega)$ make identical pricing and location decisions, so we index firms only by i and $\mathbf{z} \equiv \mathbf{z}(\omega)$. 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, or *MP cost*. For each destination n and intermediate input v, the firm chooses from among its production sites the location $\ell^\star_{in}(\boldsymbol{\varphi}(v))$ that offers the lowest destination-specific marginal cost:

$$\ell^\star_{in}(\boldsymbol{\varphi}(v)) = \arg\min_{\ell \in \mathcal{L}} \gamma_{i\ell} \frac{w_\ell}{\varphi_\ell(v)} \tau_{\ell n}, \tag{2}$$

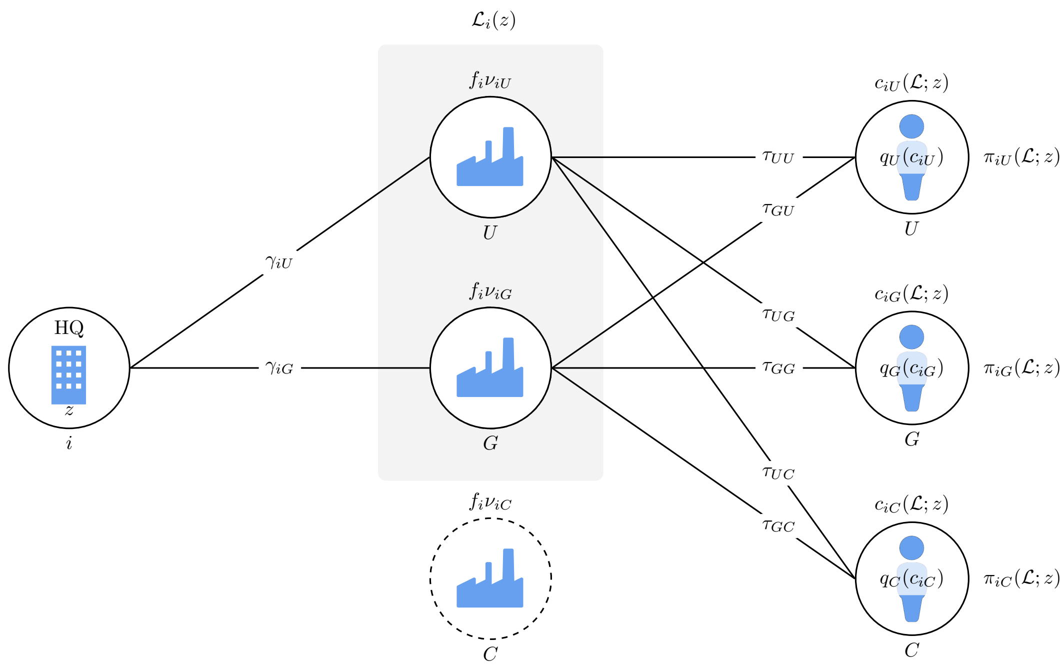
where the term $\tau_{\ell n}$ denotes a bilateral *trade costs* and the vector $\boldsymbol{\varphi}(v) = \{\varphi_\ell(v)\}_\ell$ collects a firm's productivity of producing v in every location ℓ.

We follow Tintelnot (2017) in assuming that the firm draws each of the productivity terms $\varphi_\ell(v)$ independently from a Fréchet distribution with shape θ and scale $T_\ell z_\ell$. The inverse of the shape parameter indexes the amount of productivity heterogeneity across production locations for a given variety and hence serves as a measure of the substitutability of production sites in the firm's production process. The scale parameter is the product of a location-specific productivity shifter T_ℓ, common to all firms producing in location ℓ, and the firm-specific productivity shifter z_ℓ.

The Fréchet distribution implies the following analytical expression for a firm's unit cost of delivering its final good to destination n given its headquarter location i, its productivity $\mathbf{z}$, and

[4] The demand function in equation (1) is sufficiently general to nest additively separable demand functions (as in, e.g, Krugman (1979)), the symmetric translog demand function and some of its generalizations (Feenstra (2003), Feenstra (2018)), as well as Kimball demand functions (Kimball (1995)). It also nests the demand side in canonical papers on multinational production such as Arkolakis et al. (2018) and Antras et al. (2017). In its homothetic formulation, equation (1) can be written in a spending share form and represents a class of demand functions that Matsuyama and Ushchev (2017) define as Homothetic Single Aggregator and Homothetic Direct Implicit Additivity Demands. In all these formulations, the term P_n typically represents a price aggregator and the term Q_n aggregate consumption spending.

Figure 1: A Schematic of the Model



Notes: The figure shows one possible production structure in the model for a firm headquartered in country i. The firm pays fixed costs $f_i\nu_{iU}$ and $f_i\nu_{iG}$ to operate production sites in countries U and G, so that $\mathcal{L} = \{U, G\}$. The firm incurs the MP costs γ_{iU} and γ_{iG} while producing in these countries. The production locations provide the firm's intermediate input varieties to the output markets in countries U, G, and C subject to the trade costs $\tau_{\ell U}$, $\tau_{\ell G}$, and τ_{iC}. The final assembly in the three markets relies on different varieties from both production locations, giving rise to destination-specific marginal costs.

its production location set $\mathcal{L}$:

$$c_{in}(\mathcal{L}; \mathbf{z}) = \tilde{\Gamma} \left[\sum_{\ell \in \mathcal{L}} \left(\frac{\gamma_{i\ell} w_\ell \tau_{\ell n}}{z_\ell T_\ell} \right)^{-\theta} \right]^{-\frac{1}{\theta}}, \tag{3}$$

where $\tilde{\Gamma}$ is a constant of integration and we assume $\eta < 1 + \theta$ for the marginal cost to be well-defined. A similarly tractable expression arises if, instead of the independent Fréchet distributions, the location-input-specific productivity shocks are drawn from a multivariate correlated Fréchet or Pareto distribution as in Ramondo (2014) and Arkolakis et al. (2018). The properties of the Fréchet distribution imply that the expression in equation (3) is independent of the elasticity of aggregation across varieties η (see Eaton and Kortum (2002)).

All else equal, the marginal cost in equation (3) increases in wages, MP costs, and trade costs, and decreases in firm productivity and country productivity. Crucially, the marginal cost declines in the cardinality of the production location set $\mathcal{L}$: firms that operate production sites in more locations have lower marginal costs of supplying their good to any final destination.

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. We first describe the price setting problem for a given production location set $\mathcal{L}$ and then discuss the the choice of the set itself.

Profit maximization implies that firms set the price for their final good in destination n as a markup over their marginal cost, so that:

$$p_{in}\left(\mathcal{L};\mathbf{z}\right)=\frac{\varepsilon_D\left(p_{in}\left(\mathcal{L};\mathbf{z}\right)\right)}{\varepsilon_D\left(p_{in}\left(\mathcal{L};\mathbf{z}\right)\right)-1}c_{in}\left(\mathcal{L};\mathbf{z}\right)\equiv\mu\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right)c_{in}\left(\mathcal{L};\mathbf{z}\right), \tag{4}$$

where ε_D is the price elasticity of demand, $\mu(\cdot)$ is the optimal markup as a function of marginal cost, and $p_{in}\left(\mathcal{L};\mathbf{z}\right)\equiv p\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right)$ is the price of the final good of a firm with marginal cost $c_{in}\left(\mathcal{L};\mathbf{z}\right)$ in destination market n.

Given its pricing rule, a firm's variable profit in market n is

$$\pi_{in}\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right)\equiv\left(\mu\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right)-1\right)q_{in}\left(\mathcal{L};\mathbf{z}\right)c_{in}\left(\mathcal{L};\mathbf{z}\right),$$

where $q_{in}\left(\mathcal{L};\mathbf{z}\right)\equiv q\left(p_{in}\left(\mathcal{L};\mathbf{z}\right)\right)$ is the demand for the final good of a firm with marginal cost $c_{in}\left(\mathcal{L};\mathbf{z}\right)$ in destination market n.

Finally, we turn to the choice of the production location set $\mathcal{L}$. To enter location ℓ, the firm must pay the fixed cost of opening a production location $f_i\nu_{i\ell}$. The product $f_i\nu_{i\ell}$ plays an important role in the rest of the paper and we refer to it simply as *fixed cost*. Without loss of generality, we decompose the fixed costs into a base component f_i, and a bilateral component $\nu_{i\ell}$. The firm chooses $\mathcal{L}$ to maximize the sum of destination-specific profits net of total fixed costs:

$$\mathcal{L}_i^\star(\mathbf{z})=\arg\max_{\mathcal{L}\subseteq L}\left\{\sum_{n\in N}\pi_{in}\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right)-\sum_{\ell\in\mathcal{L}}w_\ell f_i\nu_{i\ell}\right\}, \tag{5}$$

where $\mathcal{L}_i^\star(\mathbf{z})$ denotes the *optimal location decision set* of a firm headquartered in location i with productivity type $\mathbf{z}$. Note that the firm maximizes its *expected profit* when making its location decision $\mathcal{L}$ since it only learns the realization of the vector $\boldsymbol{\varphi}(v)$ *after* choosing its production location set.

Fixed Costs, Complementarities, and Combinatorial Discrete Choice In a model without fixed costs, the problem in equation (5) is trivial since $\mathcal{L}_i^\star(\mathbf{z})=L\,\forall\mathbf{z},i$, regardless of the type of complementarities among production sites. Likewise, even in the presence of fixed costs, the problem remains trivial as long as there are no complementarities among production locations. However, with *both* fixed costs and complementarities, the firm's problem becomes a combinatorial discrete choice problem, as the value of an individual production location can no longer be evaluated in isolation but now depends on which other locations the firm operates. Such CDCPs are non-trivial to solve for even moderate numbers of countries on a powerful computer. Yet, they are essential in the context of multinational production: fixed costs are necessary to account for the the empirical fact that very few, very large firms operate production sites globally. and complementarities among locations arise naturally from most model structures in the literature.

In our framework, there are two sources of such complementarities: cannibalization among production locations and scale effects. Equation (2) shows how a firm's production locations cannibalize one another's sales as they compete to be the least cost supplier, inducing a negative supply-side complementarity. The strength of the cannibalization effect 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. At the same time, the demand system induces a positive complementarity. Equation (3) shows that, all else equal, firms that have a more extensive production location set have lower marginal costs and hence larger sales volumes.; the cost savings associated with an additional production location are higher at such firms since they are applied over a larger sales volume. The strength of this complementarity 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 conclusion, both fixed costs and complementarities are hallmark features of most models of multinational production, including ours. However, their combination implies that the location choice problem in equation (5) is difficult, or even impossible, to solve. In the next section, we introduce new general methods to solve CDCPs like that in equation (5) in the presence of fixed costs and complementarities and aggregate optimal location decisions sets across heterogeneous firms to solve for market aggregates.

3 Solving and Aggregating CDCPs

We start by defining a general class of combinatorial discrete choice problems that nests the firm's profit maximization problem from our multinational production framework above.

Definition 1 (Combinatorial Discrete Choice Problem)**.** A combinatorial discrete choice problem is the problem of selecting a subset $\mathcal{L}$ of items from a finite discrete set L in order to maximize the following type of return function:

$$\pi(\mathcal{L}; \mathbf{z}) : \mathscr{P}(L) \times \mathbf{Z} \to \mathbb{R}, \tag{6}$$

where the power set $\mathscr{P}(L) = \{\mathcal{L} \mid \mathcal{L} \subseteq L\}$ is the collection of all possible subsets of L and the vector $\mathbf{z} \in \mathbf{Z} \subseteq \mathbb{R}^N$ indexes characteristics specific to the agent solving the maximization problem.

The formulation in equation (6) is very general. The return function could represent an individual's utility function, and L a set of discrete items over which the underlying preferences are defined. In many applications, other continuous variables are chosen conditional on the decision set $\mathcal{L}$, for example, production quantities chosen conditional on production locations. In such situations, equation (6) describes the CDCP that results after "maximizing out" the continuous choice variable for any given choice set. Alfaro-Urena et al. (2023) show how to embed equation (6) into a dynamic model, where the return to each potential decision set is a value function that takes into accounts its dynamic implications. In the Online Appendix , we also show that definition (1) nests the canonical Simple Plant Location Problem from the operations research literature.

While the Definition 1 is agnostic about what type of agent is solving the maximization problem, for the rest of this section, we consider a general CDCP faced by a profit maximizing firm. We refer to $\mathcal{L}$ as the firm's production location set to L as the set of countries, to $\mathscr{P}(L)$ as its choice space, and to the elements in $\mathcal{L}$ as production locations.

To formalize the interdependence among production locations, we introduce a marginal value operator that encodes the contribution to profits from a production location ℓ as part of a production location set $\mathcal{L}$.

Definition 2 (Marginal Value Operator)**.** For an item $\ell \in L$, decision set $\mathcal{L}$, and agent type $\mathbf{z}$, the marginal value operator D_ℓ on the return function $\pi(\mathcal{L};\mathbf{z})$ is defined as

$$D_\ell \pi(\mathcal{L};\mathbf{z}) \equiv \pi(\mathcal{L} \cup \{\ell\};\mathbf{z}) - \pi(\mathcal{L} \setminus \{\ell\};\mathbf{z}). \tag{7}$$

Decisions are *interdependent* and the firm's location choice problem *combinatorial* as long as the marginal value $D_\ell \pi(\mathcal{L};\mathbf{z})$, depends on the overall choice set $\mathcal{L}$. If the marginal value of any item ℓ is identical across all production location sets $\mathcal{L}$, the decision of whether to operate in location ℓ does not depend on where else the firm operates.

We also define the set-valued policy function that summarizes how the optimal decision set varies across the heterogeneous firms in the economy by mapping a firm's productivity to its optimal decision set.

Definition 3 (Policy Function)**.** Consider a CDCP as defined in Definition 1. The policy function mapping agent type to an optimal decision set is defined as the mapping $\mathcal{L}^\star : \mathbf{Z} \to \mathscr{P}(L)$ such that

$$\mathcal{L}^\star(\mathbf{z}) = \arg\max_{\mathcal{L}} \pi(\mathcal{L};\mathbf{z}).$$

The policy function is helpful for the aggregation of decisions across firms. In particular, aggregate quantities and prices are easy to compute by combining the policy function with the distribution of firm heterogeneity.

In general, a generic CDCP as defined in Definition 1 cannot be guaranteed to be solvable in less than polynomial time. However, in many economic applications, the return function πnaturally satisfies an intuitive restriction on the complementarities among production locations called single crossing differences in choices (SCD-C) that can be exploited to help solve the associated CDCP. We introduce this condition next.

3.1 Sufficient Conditions for Solving Single-Agent CDCPs

The single crossing differences property in choices is defined as follows.

Definition 4 (SCD-C)**.** Consider a return function π as defined in equation (6). For a given type $\mathbf{z}$, the return function obeys single crossing differences in choices from above if, for all $\ell \in L$ and decision sets $\mathcal{L}_1 \subset \mathcal{L}_2 \subseteq L$,

$$D_\ell \pi \left(\mathcal{L}_2; \mathbf{z}\right) \geq 0 \Rightarrow D_\ell \pi \left(\mathcal{L}_1; \mathbf{z}\right) \geq 0.$$

The return function obeys single crossing differences in choices from below if, for all $\ell \in L$ and decision sets $\mathcal{L}_1 \subset \mathcal{L}_2 \subseteq L$,

$$D_\ell \pi \left(\mathcal{L}_1; \mathbf{z}\right) \geq 0 \Rightarrow D_\ell \pi \left(\mathcal{L}_2; \mathbf{z}\right) \geq 0.$$

The single crossing restrictions are intuitive. SCD-C from above postulates that if the marginal value of setting up production in location ℓ as part of a decision set $\mathcal{L}$ is positive, it is also positive as part of any decision set $\mathcal{L}'$ where $\mathcal{L}' \subseteq \mathcal{L}$. Similarly, SCD-C from below postulates that if the marginal value of production location ℓ as part of a decision set $\mathcal{L}$ is positive, it is also positive as part of any decision set $\mathcal{L}'$ where $\mathcal{L} \subseteq \mathcal{L}'$. For the rest of the paper, we will refer to profit functions π that satisfy either SCD-C from above or below as "exhibiting SCD-C."

A simple sufficient condition for SCD-C is for the marginal value of decision ℓ, for all $\ell \in L$, to be monotone in its first argument. In particular, given any two sets $\mathcal{L}_1 \subseteq \mathcal{L}_2$, if

$$D_\ell \pi(\mathcal{L}_1; \mathbf{z}) \geq D_\ell \pi(\mathcal{L}_2; \mathbf{z}) \quad \forall \ell \in L,$$

the return function π necessarily obeys SCD-C from above and we say it satisfies the monotone substitutes property. If the weak inequality is flipped, the return function π satisfies SCD-C from below and we say it satisfies the monotone complements property.

The more restrictive monotone complements and substitute properties correspond directly to the notion of positive and negative complementarities common in economics. In particular, the marginal value of profit functions that exhibit monotone substitutes is decreasing as more locations are added to the decision set. Similarly, for profit functions exhibiting monotone complements, any element's marginal value is increasing as more locations are added to the decision set. In our setting, the definition of monotone substitutes and complements coincide with that of submodularity and supermodularity, respectively.[5]

For profit functions satisfying SCD-C, we now present a simple mapping on the firm's choice space whose fixed point corresponds to the firm's optimal set of production locations.

3.2 The Squeezing Procedure

This section presents our "squeezing procedure," a method to solve CDCPs when the underlying return function exhibits SCD-C. At the heart of the solution method is a set-valued mapping applied to the choice space associated with a CDCP. The iterative application of the mapping progressively eliminates non-optimal production location sets from the choice space and its fixed point always contains the optimal production location set. Since we consider an individual firm's problem, we do not index choice sets by firm type for notational brevity.

Consider the choice set L of the CDCP defined in equation (6). We introduce an associated pair of sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, which we call the "bounding sets." We use these sets to keep track of locations in L that are certain to be included in or excluded from the optimal set of locations $\mathcal{L}^\star$. The "subset" $\underline{\mathcal{L}}$ *includes* all locations in L which we know to *be* part of the optimal decision set. The "superset" $\overline{\mathcal{L}}$ *excludes* all items in L which we know to *not be* in the optimal decision set. The set difference of $\overline{\mathcal{L}}$ and $\underline{\mathcal{L}}$, denoted $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$, is the collection of locations which *may* be in the optimal decision set. We refer to this group as *potential* locations. A natural starting point for our procedure is to set $\underline{\mathcal{L}} = \emptyset$ and $\overline{\mathcal{L}} = L$, so that $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}} = L$, that is, so that all locations are potential production locations.

The mapping at the heart of our squeezing procedure, which we call the "squeezing step," acts on the bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ associated with the return function π. To formalize the squeezing step, we introduce an auxiliary mapping

$$\Phi(\mathcal{L}) \equiv \{\ell \in L \mid D_\ell \pi(\mathcal{L}) > 0\}$$

which collects the items $\ell \in L$ that have a positive marginal value as part of a given decision set $\mathcal{L}$. We then define the squeezing step as follows.

Definition 5 (Squeezing Step)**.** Consider a CDCP as in Definition (1) and its associated bounding sets $[\underline{\mathcal{L}}^{(k)}, \overline{\mathcal{L}}^{(k)}]$, where k indexes the output of the kth application of the squeezing step.

[5] In the Online Appendix, we show that if the choice set is not finite, the notions do not coincide. In this case, monotone substitutes and complements implies sub- and supermodularity but not vice versa.

The mapping S^a is defined as

$$S^a([\underline{\mathcal{L}}^{(k)}, \overline{\mathcal{L}}^{(k)}]) \equiv [\Phi(\overline{\mathcal{L}}^{(k)}), \Phi(\underline{\mathcal{L}}^{(k)})] \equiv [\underline{\mathcal{L}}^{(k+1)}, \overline{\mathcal{L}}^{(k+1)}]\,,$$

the mapping S^b is defined as

$$S^b([\underline{\mathcal{L}}^{(k)}, \overline{\mathcal{L}}^{(k)}]) \equiv [\Phi(\underline{\mathcal{L}}^{(k)}), \Phi(\overline{\mathcal{L}}^{(k)})] \equiv [\underline{\mathcal{L}}^{(k+1)}, \overline{\mathcal{L}}^{(k+1)}]\,.$$

If the underlying profit function π satisfies SCD-C, each application of the squeezing step adds locations to the set of locations included in the optimal set ($\underline{\mathcal{L}}$), and shrinks the set that excludes non-optimal locations ($\overline{\mathcal{L}}$). As this pair of sets updates, we eliminate some non-optimal decision sets from the choice space of the CDCP. Iteratively applying the squeezing step converges to a fixed point on the bounding sets in polynomial time. We establish both of these results in the following theorem.

Theorem 1. *Consider a CDCP as defined in equation (6).*

1. If the return function exhibits SCD-C from above, then successively applying S^a to $[\emptyset, L]$ returns a sequence of bounding sets where $\underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$.

2. If the return function exhibits SCD-C from below, then successively applying S^b to $[\emptyset, L]$ returns a sequence of bounding sets where $\underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$.

3. Conditional on the appropriate SCD-C condition, iterating on the mapping S^a or S^b converges in $O(n)$ time.

Proof. See Appendix. □

Theorem 1 ensures that every application of the squeezing step (weakly) reduces the number of potentially optimal locations.[6] In particular, the expression $\underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)}$ implies that (weakly) more locations are *included* in the subset, and hence known to be in the optimal location set, after applying the squeezing step. Similarly, the expression $\overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$ implies that (weakly) more locations are *excluded* from the superset, and hence known *not* to be in the optimal decision set, after applying the squeezing step. Crucially, no locations that are in the optimal location set are erroneously included or excluded, since $\underline{\mathcal{L}}^{(k+1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k+1)}$.

We denote the total number of iterations until convergence by K. Accordingly, we denote the operators that indicate applying the mappings S^a and S^b until convergence by $S^{a(K)}$ and $S^{b(K)}$, and the resulting bounding sets by $[\underline{\mathcal{L}}^{(K)}, \overline{\mathcal{L}}^{(K)}]$. If the converged pair of bounding sets

[6] The squeezing step is designed to recover $\mathcal{L}^\star$ so that all potential locations ℓ for which the firm is indifferent are excluded from the optimal strategy. In applications in which these potential locations should be included as part of the optimal strategy, they are easily identified as locations ℓ for which $D_\ell \pi(\mathcal{L}^\star) = 0$.

is identical such that $\underline{\mathcal{L}}^{(K)} = \overline{\mathcal{L}}^{(K)}$, Theorem 1 implies that $\mathcal{L}^{\star} = \underline{\mathcal{L}}^{(K)} = \overline{\mathcal{L}}^{(K)}$ so that we have identified the optimal set of locations that maximizes the underlying profit function of the firm.[7]

There are cases when the converged pair of bounding sets is not identical, so that there are several *potentially* optimal production locations left. One option is then to compare the value of profits for all remaining potentially optimal production location sets ("brute force" approach). Since the brute force approach is computationally expensive, we introduce a refinement which we refer to as "branching procedure" and formally characterize it in Online Appendix OA.4.1. The branching procedure often finds the optimal location set much more quickly, and only collapses to the brute force method in the worst-case scenario.

At the heart of the solution method for CDCPs with supermodular profit functions outlined in Jia (2008) is an increasing set-valued mapping whose set of fixed points always contains the profit-maximizing location set as shown in Tarski (1955). If the underlying profit function is submodular (i.e., exhibits negative complementarities), however, the theorem in Tarski (1955) does not apply. In Online Appendix OA.7, we rewrite our results in terms of lattice algebra to make them directly comparable to those in Jia (2008) and provide an in-depth comparison of the two papers.

3.3 Sufficient Conditions for Solving for the Policy Function

In this section, we show how to solve for the policy function mapping firm types into optimal decision sets in settings in which heterogeneous firms each solve a CDCP. To that end, we introduce an additional restriction on the return function called single crossing differences in type (SCD-T) that has implications on how the optimal location set changes with agent type.

We define the set of all firm types $\mathbf{z}$ that derive a positive marginal value from the inclusion of location ℓ as part of a production location set $\mathcal{L}$:

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

Using this set, we introduce the follow restriction on the return function π:

Definition 6 (SCD-T)**.** The return function π exhibits single crossing differences in type if, for all items ℓ and decision sets $\mathcal{L}$, $\Lambda_{\ell}(\mathcal{L})$ and its complement $\Lambda^{c}_{\ell}(\mathcal{L})$ are both connected sets.

The two contiguous sets $\Lambda_{\ell}(\mathcal{L})$ and $\Lambda^{c}_{\ell}(\mathcal{L})$ divide the type space $\mathbf{Z}$ into types which receive positive marginal value and types which receive negative marginal value from location ℓ's

[7] In the Appendix, we establish that K is never larger than the cardinality of the choice set, $|L|$. Note that our approach requires that the return function satisfies the same type of SCD-C (i.e., either below or above) over the entire choice space. For a given set of parameters, the structure of economic models typically implies that the return function exhibits the same type of SCD-C over the entire choice space.

inclusion in the set of locations $\mathcal{L}$.[8]

Intuitively, the SCD-T condition postulates that if a production location ℓ has a positive marginal value for a firm of type $\mathbf{z}$ as part of a production location set $\mathcal{L}$, firms with sufficiently similar type $\mathbf{z}'$ also derive a positive marginal value from producing in location ℓ. SCD-T restricts firms with similar productivity to have the same optimal decision set.

A sufficient condition for SCD-T follows from super- and sub-modularity between type and decision set. In particular, fix a location ℓ, set $\mathcal{L}$, and component j of the multi-dimensional type vector. Holding all other components of the type vector constant, one can check if the selected component j exhibits either supermodularity or submodularity with the decision set. If it does for every possible item ℓ, set $\mathcal{L}$, and component j, then the return function exhibits SCD-T. Note that the sufficient condition allows for a given component j to be supermodular with item and decision set $(\ell, \mathcal{L})$, but submodular with different pair $(\ell', \mathcal{L})$. Likewise, it allows for a given component j to be supermodular with a pair $(\ell, \mathcal{L})$, while another component ℓ' is submodular with the same pair.

For illustration, Figure 2 depicts a policy function associated with a one-dimensional type space obeying these restrictions. In the figure, firms with types between z_1 and z_2 all have an optimal production location set $\mathcal{L}_1^\star$, while firms between z_2 and z_3 have the optimal strategy $\mathcal{L}_2^\star$, and so on. The SCD-T assumption ensures that the set-valued policy function changes only changes value at discrete points in the type space, e.g., z_1 and z_2.

In the most prominent case of one-dimensional firm heterogeneity, we can write the SCD-T restriction in parallel with the SCD-C restriction in Section 3.1. In particular, given two types $z_1 < z_2$, SCD-T asserts, for all elements $\ell \in L$ and decision sets $\mathcal{L}$,

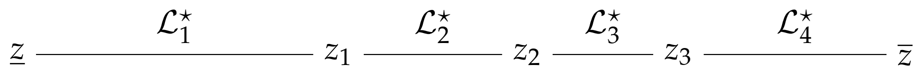
$$D_\ell \pi(\mathcal{L}; z_1) \geq 0 \Rightarrow D_\ell \pi(\mathcal{L}; z_2) \geq 0.$$

Furthermore, in the case of one-dimensional firm heterogeneity, if the return function obeys both SCD-C from below and SCD-T, the resulting policy function obeys a *nesting* structure. That is, given two scalar types $z_1 < z_2$, it must be the case that $\mathcal{L}^\star(z_1) \subseteq \mathcal{L}^\star(z_2)$, as shown in Antras et al. (2017) for supermodular objective functions.[9] We generalize this nesting result to a multidimensional type space under a stronger restriction in place of SCD-T in the Online Appendix. Importantly, with SCD-C from *above* instead of *below* the policy function does not necessarily obey a nesting structure. In the context of the location problem, there is no strict "pecking order of production locations." Instead, more productive firms may choose sets that contain fewer and different production locations than less productive firms, and vice versa.

[8] The SCD-T property is therefore implied by the supermodularity property introduced by Costinot (2009) (in its log form).

[9] Topkis (1978) shows that the policy function exhibits a nesting structure in settings with positive complementarities, a single dimension of agent heterogeneity, and supermodularity of agent type with choices, in which more productive types have optimal choice sets that nest those of less productive types.

Figure 2: The Policy Function Over a One-Dimensional Type Space



Notes: The figure shows the one dimensional type space $[\underline{z}, \overline{z}]$ on a line. For illustration, it also shows groups of agent types that have the same optimal location set. The single crossing differences in types (SCD-T) assumption ensures that such groups exist. The resulting policy function $\mathcal{L}^\star(\cdot)$ changes value only at each cutoff z_n, for $n \in \{1, 2, 3\}$.

The resulting optimal policy function is a complicated object that is difficult to theoretically characterize. This challenge motivates our all-inclusive solution approach, which does not rely on any particular property of the policy function, and only requires SCD-C and SCD-T to hold for the underlying return function.

Figure 2 illustrates that recovering the policy function requires solving for *both* its "kink points" $\{z_1, z_2, z_3\}$, and the optimal decision sets for the subregions of types they create. In the next section, we introduce a "generalized squeezing procedure" that simultaneously solves for both.

3.4 The Generalized Squeezing Procedure

With heterogeneous firms, we extend the notion of bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, associated with a CDCP, to set-valued functions over the type space, $\underline{\mathcal{L}}(\cdot)$ and $\overline{\mathcal{L}}(\cdot)$. These "boundary set functions" are such that $\underline{\mathcal{L}}(\mathbf{z}) \subseteq \mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}(\mathbf{z})$ for any type $\mathbf{z}$.

With these concepts in hand, we introduce a "generalized squeezing step." The squeezing step in Section 3.2 acted on the boundary sets associated with the CDCP. The generalized squeezing step instead acts on a region of the type space $Z \subseteq \mathbf{Z}$ such that for all $\mathbf{z} \in Z$ the current boundary set functions map to the same subset $\underline{\mathcal{L}}$, and superset $\bar{\mathcal{L}}$, for all $z \in Z$. We collect all information on a region Z relevant to the squeezing step in a 4-tuple, $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]$, where $\underline{\mathcal{L}}$ and $\overline{\mathcal{L}}$ are the the current bounding sets over the interval Z, that is, the values of the bounding set functions evaluated for any $z \in Z$. The "auxiliary" set M collects potential production locations the algorithm has already considered but could not make progress on.

Whereas an application of the squeezing step from Section 3.2 updates only the boundary sets, an application of the *generalized* squeezing step updates *both* the boundary sets and refines the partition of the type space for which current boundary sets are identical (i.e., adds new "kinks" to the boundary set functions). In particular, applying the generalized squeezing step to a given 4-tuple creates up to three new 4-tuples, each corresponding to a subregion of the original region Z, and each with either updated boundary sets or an updated auxiliary set. The technique is recursive, since the generalized squeezing step creates several 4-tuples from an initial 4-tuple at each application. The eventual output is a collection of 4-tuples each with associated boundary sets. As with the simple squeezing procedure, ideally the boundary sets of each 4-tuple coincide with one another in which case they also coincide with the optimal strategy for all types in the associated subregion of the type space Z.

We now define the "generalized squeezing step," which when applied to a 4-tuple creates up to three new 4-tuples each defined over a subregion of the type space for which the original 4-tuple was defined.

Definition 7 (Generalized Squeezing Step)**.** Consider a CDCP faced by agents on a type space $\mathbf{Z}$, and a subregion of its type space $Z \subseteq \mathbf{Z}$ with associated bounding sets $(\underline{\mathcal{L}}, \overline{\mathcal{L}})$ and auxiliary set M. Summarize it by the 4-tuple $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]$ and select some element $\ell \in \overline{\mathcal{L}} \setminus (M \cup \underline{\mathcal{L}})$.

The mapping S^a is defined as

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

where any 4-tuple with empty subregion may be omitted.

The mapping S^b is defined as

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

where any 4-tuple with empty subregion may be omitted.

The next theorem establishes that if a CDCP's underlying return function exhibits SCD-C and SCD-T, then each application of the generalized squeezing step updates the bounding sets without excluding locations that are part of the optimal decision for any subregion Z of the type space.

Theorem 2. *Consider a CDCP as defined in equation (6) for agents on a type space* $\mathbf{Z}$*, and associated 4-tuple* $[(\underline{\mathcal{L}}_0, \overline{\mathcal{L}}_0, M), Z]$ *for which* $M \subseteq (\overline{\mathcal{L}}_0 \setminus \underline{\mathcal{L}}_0)$ *and* $\overline{\mathcal{L}}_0 \subseteq \mathcal{L}^\star(\mathbf{z}) \subseteq \underline{\mathcal{L}}_0$ *for all* $\mathbf{z} \in Z$*. Suppose the underlying return function exhibits SCD-T over* Z*.*

1. If the underlying return function π exhibits SCD-C from above, then applying the mapping S^a recursively partitions Z into disjoint subregions. Further, $\underline{\mathcal{L}}_0 \subseteq \underline{\mathcal{L}}(\mathbf{z}) \subseteq \mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}(\mathbf{z}) \subseteq \overline{\mathcal{L}}_0$ for each $\mathbf{z} \in Z$.

2. If the underlying return function π exhibits SCD-C from below, then applying the mapping S^b recursively partitions Z into disjoint subregions. Further, $\underline{\mathcal{L}}_0 \subseteq \underline{\mathcal{L}}(\mathbf{z}) \subseteq \mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}(\mathbf{z}) \subseteq \overline{\mathcal{L}}_0$ for each $\mathbf{z} \in Z$.

3. Conditional on the appropriate SCD-C restriction, the recursive application of mappings S^a and S^b converges in $O(n)$ time.

Given a CDCP with underlying return function exhibiting SCD-C and SCD-T, we can define the generalized squeezing procedure as recursively applying the generalized squeezing step until $\overline{\mathcal{L}} = M \cup \underline{\mathcal{L}}$ on each subregion of the type space. The squeezing procedure has converged globally when it converges on all 4-tuples separately. Given Theorem 2, the generalized squeezing procedure delivers bounding set functions $\underline{\mathcal{L}}(\cdot)$ and $\overline{\mathcal{L}}(\cdot)$ such that $\underline{\mathcal{L}}(\mathbf{z}) \subseteq \mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}(\mathbf{z})$ for all $\mathbf{z} \in \mathbf{Z}$. For subregions where $\underline{\mathcal{L}}(\mathbf{z}) \subset \overline{\mathcal{L}}(\mathbf{z})$ for some type $\mathbf{z}$ after global convergence of the generalized squeezing procedure, we provide a refinement of the brute force approach that we refer to as generalized branching procedure in the Online Appendix.

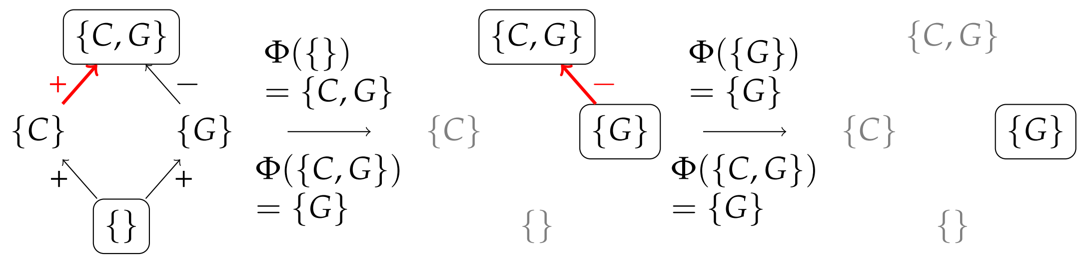
A Simple Example To elucidate our solution methods, we consider a stylized version of the plant location decision in our theoretical framework in Section 2 with only two available countries, Germany (G) and Canada (C). The firm's choice set is $L = \{G, C\}$ and its choice space is $\mathscr{P}(L) = \{\{\}, \{G\}, \{C\}, \{G, C\}\}$, which contains all potential decision vectors of the firm. Suppose the firm's return function satisfies SCD-C from above.

Figure 3 depicts the steps of applying the squeezing method to the firm's location choice problem. The signs along the arrows indicate the marginal value of adding a given location to the bounding set. Consider the leftmost panel. The initial bounding pair is $[\{\}, \{G, C\}]$ so that $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}$, as required. Beginning with the subset, $\underline{\mathcal{L}} = \{\}$, we consider the marginal value of Canada and Germany, separately, given the empty decision set. Since both countries have positive marginal values and SCD-C from above holds, we cannot discard either location as suboptimal. Next, we evaluate the respective marginal value of including Canada and Germany in the superset, $\overline{\mathcal{L}} = \{G, C\}$. Given SCD-C from above, Germany's marginal value remains positive when Canada is present. The optimal location set hence includes Germany with certainty. We can draw no inference about Canada's inclusion in the optimal decision set. This completes the first application of the squeezing step.The updated bounding sets in the middle panel of Figure 3 reflect that Germany is included in the optimal location set with certainty. Since the marginal value of Canada is negative in this context, we conclude that the firm optimally only operates in Germany, so that $\mathcal{L}^\star = \{G\}$.

Figure 4 illustrates an application of the generalized squeezing step to solve for the policy function in the same context. As before, the signs along the arrows indicate the marginal value of adding a given location to the bounding set, though now this marginal value may vary for different firm type. Red arrows imply that an interval's bounding pair can be updated.

We begin at the natural starting point, with the 4-tuple $[\{\}, \{C, G\}, \{\}, (-\infty, \infty)]$, which represents the entire productivity range with the bounding sets $\{\}$ from below and $\{C, G\}$ from above. In the figure, we consider adding a Germany as a production location. We identify two cutoff productivities, marked with vertical ticks. For firm types below the first cutoff z_G^{out}, adding Germany when Canada is excluded yields negative marginal benefit; these firms therefore never operate in Germany. This subinterval is $\Lambda_G^c(\{\})$, for which we update the superset by omitting

Figure 3: An Example Application of the Squeezing Step



Notes: Iteratively applying the squeezing step twice to an example problem with SCD-C from below. The choice space shrinks with each application. In every step, the current subset and superset are circled. Arrows indicate whether profits increase or decrease when switching from one decision set to another, where red arrows indicate an update is possible. In the first step, the subset is the empty set and the superset is the full set of locations. Each location is valuable if the other is not present, as indicated by the positive change from $\{\}$ to either $\{C\}$ or $\{G\}$. Thus, $\Phi(\{\}) = \{C, G\}$, and we cannot yet optimally exclude either location. However, it is still valuable to add the G location even if the firm is active in the Canadian location, as indicated by the red arrrow. Thus, $\Phi(\{C, G\}) = \{G\}$ and we optimally include the German location, adding it to the subset. The middle panel illustrates the problem after one update from the squeezing step. Applying the squeezing step again, we use the fact that adding the Canadian location to the subset, which now includes the German location, is suboptimal. Thus, $\Phi(\{G\}) = \{G\}$, and we optimally exclude the Canadian location from the superset. Ultimately, one decision set remains.

Germany. For firm types above the second cutoff z_G^{in}, operating in Germany while also operating in Canada yields positive marginal benefit; these firms therefore always operate in Germany. This subinterval is $\Lambda_G(\{C, G\})$, for which we update the subset by adding Germany. For firms in the middle interval, no update is possible, so G enters the auxiliary set M of "undetermined" locations. The generalized squeezing step hence divides original productivity range into three new 4-tuples. To continue the algorithm, we recursively apply the generalized squeezing step to each new 4-tuple, continuing until each 4-tuple has converged.

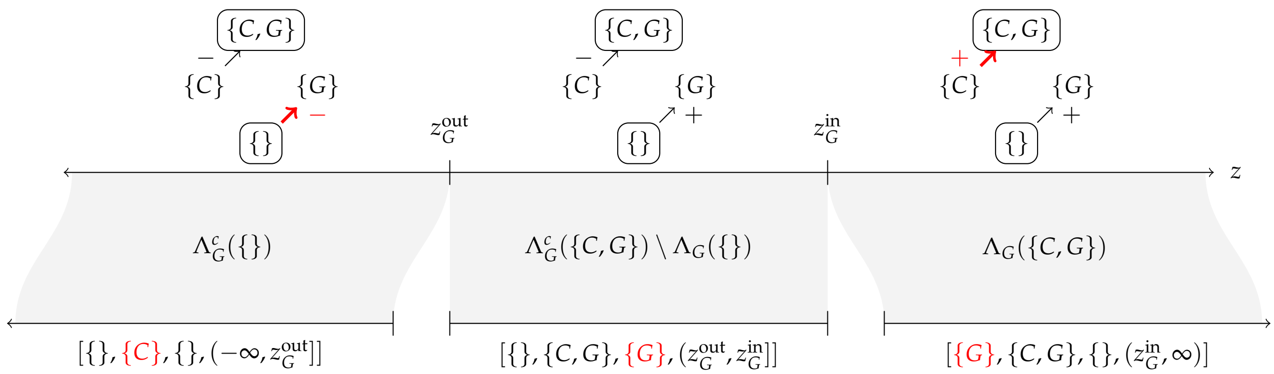
4 Establishing SCD and Closing the Model

In this section, we establish the SCD-C and SCD-T conditions in our model of multinational production from Section 2. We then introduce additional parametric assumptions in preparation for the quantitative application below and specify the full set of equilibrium conditions. For the rest of the paper, we assume that the firm-specific productivity shifter does not differ across locations, so that $z_\ell = z \ \forall \ell$. This allows us to provide sharper intuition and is the relevant case in most quantitative applications.

4.1 SCD-C and SCD-T in our Quantitative Framework

Consider the framework presented in Section 2. We show in the Online Appendix, that the profit maximization problem of a firm headquartered in location i with productivity z satisfies

Figure 4: An Example Application of the Generalized Squeezing Procedure



Notes: The figure shows an example single application of the generalized squeezing procedure. The problem in the figure satisfies SCD-C from above and the type space is single dimensional. We initialize the procedure with a natural initial 4-tuple, $[\{\},\{C,G\},\{\},(-\infty,\infty)]$. We consider whether firms optimally operate in Germany. This application divides the original type space into three subintervals, the leftmost of which we can discard Germany from consideration, the rightmost of which we can definitely include Germany, and the middle of which we cannot yet make a decision for the Germany as a production location. This application leaves three 4-tuples, to each of which we can apply the generalized squeezing step again.

SCD-C from below as long as

$$\underbrace{\frac{\mathrm{d}\ln q_{in}(\mathcal{L};z)}{\mathrm{d}\ln p_{in}(\mathcal{L};z)}\frac{\mathrm{d}\ln p_{in}(\mathcal{L};z)}{\mathrm{d}\ln c_{in}(\mathcal{L};z)}}_{\text{Demand Channel}} \geq \underbrace{1+\theta}_{\text{Supply Channel}} \tag{8}$$

for all destinations n and decision sets $\mathcal{L} \subseteq L$. SCD-C from above holds if the inequality is reversed.

The expression for the marginal cost of a firm (cf. equation 3) shows that an additional production location always lowers the marginal cost of the firm to supply its final good to any location. The firm's problem exhibits positive complementarities when an additional locations leads to a larger profit gain the more locations the firm operates, and vice versa for negative complementarities. Equation (8) decomposes this effect into a supply-side component and a demand-side component. The supply-side component captures how much an additional production location reduces the marginal cost of the firm, while the demand-side component captures by how much variable profits increase when the marginal cost of the firm drops. The balance of these two forces determines whether the firm's overall profit maximization problem exhibits positive or negative complementarities.

The strength of the supply-side channel depends on how substitutable production locations are in the production process of the firm, which is parameterized by the dispersion ($1/\theta$) of location-input-specific productivity shocks. If $1/\theta$ is low, the firm selects its least cost production locations mainly based on trade costs and productivities common across all varieties produced in the same location, which leaves locations more substitutable. In this case, cannibalization among production sites is stronger. At the same time, the strength of the demand-side channel depends on the product of the demand and passthrough elasticity. It summarizes the elasticity of variable profits to a change in marginal cost, which is determined by how much a marginal cost change affects the price (passthrough) and in turn by how much demand responds to a marginal decrease in price (demand elasticity).

The product of the demand and pass-through elasticities also plays a central role in establishing SCD-T. In particular, SCD-T holds as long as

$$\frac{\mathrm{d}\ln q_{in}(\mathcal{L};z)}{\mathrm{d}\ln p_{in}(\mathcal{L};z)}\frac{\mathrm{d}\ln p_{in}(\mathcal{L};z)}{\mathrm{d}\ln c_{in}(\mathcal{L};z)} \geq 1. \tag{9}$$

Intuitively, the SCD-T condition requires that the variable profit increase associated with an additional production location is higher at more productive firms, akin to a cross-derivative. The condition in equation (9) again separates into a demand-side effect on the left and a supply-side effect on the right. The demand-side effect is identical to the one from SCD-C, and describes by the elasticity of variable profits to marginal costs. The supply-side effect captures how the reduction in marginal costs associated with an additional production location interacts with firm productivity. An additional production location is worth more at an unproductive firm compared to a productive firm, since the productive firm has high marginal costs but can shore up its low productivity by establishing more production locations. In other words, productivity and production sites are substitutes in the firm's cost function. As productivity enters the cost function multiplicatively (see equation 3), the elasticity of substitution between the benefit of a production location and the firm's innate productivity is simply 1.

In the Online Appendix OA.3, we provide the proofs for the results in this section and 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 Section 2 is a special case. There, we also show how to establish SCD-C and SCD-T for models with location-input-specific productivity shocks are drawn from a multivariate Fréchet distribution or a multivariate Pareto distribution as in Ramondo (2014) and Arkolakis et al. (2018) or from distributions with correlated draws as in Lind and Ramondo (2023) and Xiang (2022).

4.2 The Equilibrium System in the Quantitative Framework

For our quantitative exercise, we specialize the general demand function in equation (1) to that implied by a CES demand system with elasticity σ, as is common in the multinational production literature (Helpman et al. (2004); Tintelnot (2017); Arkolakis et al. (2018)). The resulting pricing rule is a special case of the rule in equation (4) with a *constant* markup over marginal cost that is independent of the firm's production location set. In the Online Appendix, we also present a version of the model with the Pollak (1971) demand system , which generates *variable* markups. Under the CES assumption, equation (4) collapses to $\sigma \geq 1+\theta$, which reflects that markups are constant, pass-through is 1, and the demand elasticity is simply σ. The condition for SCD-T collapses to $\sigma \geq 1$. Therefore, in the CES case, SCD-C from below implies SCD-T.

We finally turn to aggregation over firms and the determination of aggregate variables in general equilibrium. We make use of the policy function notation for location choices in the equilibrium conditions. In particular, We denote by $\mathcal{L}_i^\star(z)$ the production location set of a firm with productivity z headquartered in location i.

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 the distribution $G_i(z)$. After learning their productivity draw, firms decide whether to incur the fixed cost $f_i \nu_{i\ell}$ of setting up production in location ℓ. Given these fixed costs, entrants with productivity below a cutoff productivity $\tilde{z}_i$ do not find it profitable to operate any production location and hence never produce positive quantities. The cutoff $\tilde{z}_i$ is pinned down by the zero profit condition:

$$\pi_i(\mathcal{L}_i^\star(\tilde{z}_i);\tilde{z}_i) = 0. \tag{10}$$

Firms enter until expected profits are zero. As a result, the total mass of entrants M_i in each origin country i satisfies the free entry condition

$$w_i f_i^e = \frac{1}{\sigma}\sum_n X_n \int_{\tilde{z}_i}^{\infty} \left(\frac{p_{in}\left(\mathcal{L}_i^\star(z);z\right)}{P_n}\right)^{1-\sigma} \mathrm{d}G_i(z) - \sum_\ell w_\ell f_i \nu_{i\ell} \int_{\tilde{z}_i}^{\infty} \mathbb{1}_{i\ell}(z)\mathrm{d}G_i(z) \tag{11}$$

where $\mathbb{1}_{i\ell}(z)$ indicates whether a firm with productivity z from origin i optimally opens a production location in country ℓ, and X_n denotes total expenditure on final goods in destination n. The mass of final product varieties offered by firms headquartered in location i is then simply $(1-G_i(\tilde{z}_i))M_i$.

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

$$P_n^{1-\sigma} = \sum_i M_i \int_{\tilde{z}_i}^{\infty} p_{in}\left(\mathcal{L}_i^\star(z);z\right)^{1-\sigma} \mathrm{d}G_i(z). \tag{12}$$

Next, the labor market must clear in each production location ℓ.[10] There are three sources of labor demand: variable labor requirements from all the production sites operating in country ℓ,

[10]To characterize the total labor demand from variable production in a specific production country ℓ, we use the fact that they are a fixed proportion $(\sigma-1)/\sigma$ of total sales under CES demand. Of these total production costs, production location ℓ accounts for the share $\frac{(\gamma_{i\ell} w_\ell \tau_{\ell n}/T_\ell)^{-\theta}}{\sum_{\ell' \in \mathcal{L}_i^\star(z)}\left(\gamma_{i\ell'} w_{\ell'} \tau_{\ell' n}/T_{\ell'}\right)^{-\theta}}$ out of the firm's complete production location set $\mathcal{L}_i^\star(z)$.

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.

$$w_\ell H_\ell = \frac{\sigma-1}{\sigma}\sum_{i,n} X_n M_i \int_{\tilde{z}_i}^{\infty} \frac{\mathbb{1}_{i\ell}(z)\left(\gamma_{i\ell} w_\ell \tau_{\ell n}/T_\ell\right)^{-\theta}}{\sum_{\ell'\in\mathcal{L}_i^\star(z(\omega))}\left(\gamma_{i\ell'} w_{\ell'}\tau_{\ell' n}/T_{\ell'}\right)^{-\theta}}\left[\frac{p_{in}\left(\mathcal{L}_i^\star(z);z\right)}{P_n}\right]^{1-\sigma} \mathrm{d}G_i(z) \qquad (13)$$
$$+ M_\ell w_\ell f_\ell^e + \sum_i w_\ell f_i \nu_{i\ell} M_i \int_{\tilde{z}_i}^{\infty} \mathbb{1}_{i\ell}(z)\mathrm{d}G_i(z).$$

Finally, balance of payments implies that total expenditure from consumers in a market n must equal their total income:

$$X_n = w_n H_n \qquad (14)$$

A general equilibrium in our model is defined as follows:

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

1. given the aggregates, the policy functions solve the firm's optimization problem in equation (5); and

2. given the policy functions, the aggregates satisfy equations (10), (11), (12), (13), and (14).

For the quantitative implementation, we assume that the firm productivity distribution $G_i(z)$ is Pareto with shape parameter ξ and country of headquarter-specific Pareto minimum $\underline{z}_i$.

5 Quantification

In this section, we calibrate our theoretical framework to prepare it for counterfactual analysis. We first describe our data sources, then our calibration strategy, then finally the fit of the calibration. We use the calibrated model to illustrate the performance of our squeezing and branching steps in several speed tests. Appendix B provides additional information on data sources, calibration, and model fit.

5.1 Data

We obtain information on manufacturing trade and multinational production for 32 countries from Alviarez (2019).[11] For each host country, the data set contains the number of manufacturing

[11] 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,

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 trade flows for all country pairs. For all variables, the data reflect average values over the period from 2003 to 2012.

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.

We use the standard bilateral gravity variables from the CEPII database (see Conte et al. (2023)): distance, a shared borders dummy, an indicator for past colonial relationships, and a common language dummy. In addition, we draw on information from the TRAINS dataset to construct several measures of bilateral manufacturing tariffs, closely following the existing literature. We also add information on real GDP per capita and total employment for each country from the Penn World Tables (Feenstra et al. (2015)).

5.2 Calibration Strategy

In this section, we outline our calibration strategy. We set the elasticities governing the strength and type of complementarities among production locations to canonical values in the literature. We parameterize all bilateral cost matrices as constant elasticity functions of standard gravity variables, then infer these elasticities simultaneously with the productivity parameters using a method of moments estimator. Table 2 summarizes the calibrated parameters that are not country-specific.

Elasticity of Substitution and Dispersion of Productivity Shocks: σ and $1/\theta$ The elasticity of substitution σ across final consumption goods, and the dispersion $(1/\theta)$ of location-input-specific productivity shocks, together determine the direction of the complementarities among production locations. In our baseline calibration, we follow Arkolakis et al. (2018), to set $\sigma = 4$ and $\theta = 4.5$. Together, our choices for σ and θ imply *negative* complementarities among a firm's production locations since $\sigma < \theta + 1$.

Our choice of σ falls within the range of estimates of Broda and Weinstein (2006). Arkolakis et al. (2018) show that it is also consistent with markup estimates from the manufacturing sector. Moreover, our value is similar to the estimate of $\sigma = 3.89$ from Head and Mayer (2019). While their study 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

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.

Table 1: Trade, MP , and Foreign Affiliates Gravity in the Data

	Trade (1)	MP (2)	Affiliates (3)
Log Distance	-0.687***	-0.289**	-0.681***
	(0.0541)	(0.106)	(0.0847)
Colony	0.0776	-0.00471	0.232
	(0.125)	(0.131)	(0.144)
Contiguity	0.448***	0.424**	0.448***
	(0.0697)	(0.164)	(0.0979)
Language	0.152	0.468**	0.561***
	(0.102)	(0.145)	(0.155)
Observations	992	992	992

Notes: The table presents the estimated coefficients from estimating a standard log linear gravity equation using Poisson Pseudo Maximum Likelihood. The outcome variable differs across the three columns: bilateral manufacturing trade flows (Column 1), bilateral multinational production sales (Column 2), 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. 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.

estimates from the car industry in Goldberg (1995) and Berry et al. (1995).

Our choice of θ generates trade elasticities ranging from 2.8 to 5, depending the specification. 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. Head and Mayer (2019) provide the main alternative estimate of θ in the literature of multinational production; they estimate $\theta = 7.7$. For robustness, we calibrate a version of our model the value for θ from which implies even stronger negative complementarities than our baseline of $\theta = 4.5$. Finally, for illustration, we also provide an alternative calibration that uses $\theta = 2.5$, which implies *positive* complementarities. Online Appendix OA.8 contains the details of these alternative calibrations.

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 component of fixed costs $\{\nu_{i\ell}\}_{i\ell}$. We follow Tintelnot (2017) and parameterize these three matrices as constant elasticity functions of the standard bilateral gravity variables: 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. In addition, we include a destination-specific fixed effect for activity (exporting, producing, and setting up a production location) abroad versus at home

Table 2: Central Parameters in the Calibrated Model

	Parameter	Value
σ	Elasticity of Substitution across final goods	4
θ	Elasticity of Substitution among production locations	4.5
η	Elasticity of Substitution among inputs	3
ξ	Dispersion of Firm productivity distribution	4.95

Notes: The tables shows the values for the main parameters in the model.

in all bilateral costs. For the trade costs, we also include tariffs as observed in the data which enter the costs with an elasticity of one.[12] We present the expressions for each element of the bilateral cost matrices in Appendix B.2.2.

Table 1 presents the results from estimating gravity equations for trade flows, inward MP sales, and inward affiliate stocks using a Poisson Maximum Likelihood (PPML) estimator. All specifications in Table 1 include origin and destination fixed effects, as well as additional controls for tariffs and trade agreements. In Appendix B.2.1, we present the formal PPML specification and a variety of robustness checks on the estimated coefficients. We also present an OLS version of Table 1 where the estimated elasticities are in line with those of similar regressions in Ramondo et al. (2015).

We choose the elasticities of the various gravity variables in the bilateral cost matrices to ensure the model-generated data produces estimated coefficients similar to those in Table 1. In particular, we employ an indirect inference approach which involves solving the full model many times with different guesses for the elasticities. We choose the three sets of destination fixed effects for activity at home versus abroad to exactly match the own shares of trade, the total sales of domestic firms, and the total stock of domestic firms in each country observed in the data.

Country Productivity Parameters and Entry Costs The model features two vectors of country-specific productivity terms: the scale parameter $\underline{z}_i$ of the firm Pareto distribution, which acts as a location-of-headquarter productivity shifter, and the scale parameter T_ℓ of the Fréchet distribution of location-input-specific productivity shocks, which acts as a location-of-production productivity shifter. We choose $\underline{z}_i$ to exactly match the total amount of foreign affiliate sales conducted by firms headquartered in i relative to the US, and T_ℓ to exactly match the observed GDP per capita for each country relative to the US.

Before making production decisions, the firms in our model must incur an entry cost $f_i^e > 0$ to draw a productivity. Next, they decide whether to incur the fixed cost $f_\ell \nu_{i\ell}$ to establish a

[12]For any bilateral pair where there is no MP activity in the data, we set the MP costs to infinite: $\gamma_{i\ell} = \infty$.

production site in one or more locations ℓ and commence production. Given these two costs, only a fraction of firms produce in positive quantities after entry. For a given entry cost f_i^e, we choose the base component of fixed costs f_ℓ to match the share of "surviving" firms whose productivity draw exceeds the cutoff level defined by the zero profit condition in equation (10). In a second step, we recover the entry cost f_i^e by inverting the free entry condition in equation (11) so that the model matches the number of entering firms observed in the data.

Other Parameters The Pareto distribution of firm productivity is governed by the dispersion parameter ξ. The model generates a closed form expression for the Pareto tail of the firm sales distribution, which is $\xi/(\sigma-1)$. We choose ξ to match the estimates presented in Arkolakis (2010) and set $\xi = 1.65 \times (\sigma - 1)$. The parameter η governs the elasticity of substitution among a firm's own inputs. As in Antras et al. (2017), this elasticity does not affect general equilibrium allocations, so we set it to 3 without loss of generality.[13] We take the number of workers in each country directly from the data.

Alternative Calibration without Fixed Costs To better understand the role of fixed costs in determining economic allocations, we also calibrate a version of the model where we set the fixed costs to zero, i.e., we set $f_i \nu_{i\ell} = 0\,\forall i$. For this calibration, we follow the same procedure as above but drop the coefficients in the third column of Table 1 as calibration targets. Without fixed costs, firms open production locations everywhere and the firm faces a degenerate combinatorial problem. We compare results from this calibration to our baseline calibration to understand the importance of fixed costs in shaping counterfactual responses of the economy to economic shocks, and the interaction between complementarities and fixed costs. We provide details and all results of this alternative calibration in Online Appendix OA.8.2.

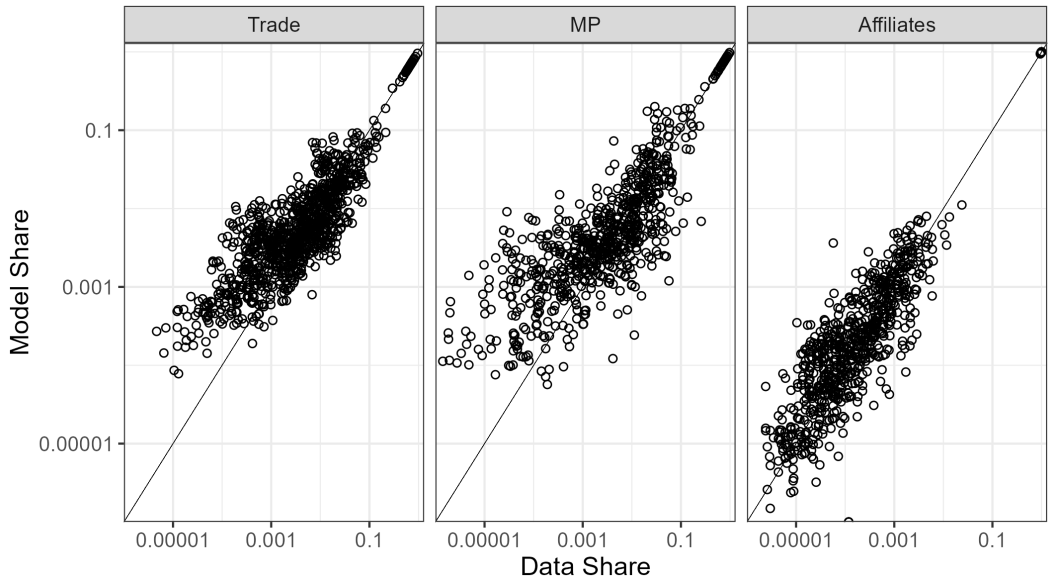
5.3 Model Fit

Figure 5 graphs the trade shares, inward MP sales shares, and inward foreign affiliate shares generated by the calibrated model against the same objects in the data. The model provides a good fit, especially for larger shares. The fit for larger shares is better since the targeted PPML specification in Table 1 is in levels, thereby putting relatively more weight on larger countries (see Sotelo (2019) for a discussion). The model-generated data produces exactly the same coefficient estimates as in Table 1 (see Table OA.1 in Appendix OA.2).

Figure 6 shows a histogram 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

[13] Note that the model requires $\eta \in (1, \theta + 1)$. Since our lowest θ is equal to 2.5, the choice of $\eta = 3$ ensures that this parameter restriction is always satisfied.

Figure 5: Trade Shares, Inward MP Sales Shares, and Inward Affiliate Shares in the Data and the Baseline Calibration



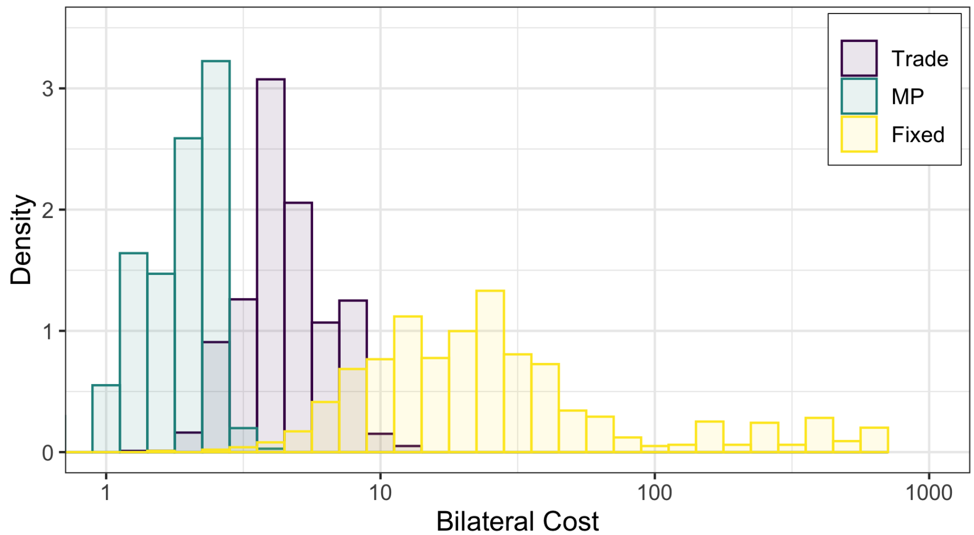
Notes: The figures shows graphs statistics from the data obtained from Alviarez (2019) against the same objects in the calibrated model. The left panel shows trade shares, the second panel shows inward MP sales shares, and the third panel shows inward MP affiliate shares. The correlations between the off-diagonal shares in the model and data are 0.757, 0.726, and 0.824 are respectively.

(2013)).

In contrast, our estimated MP costs are small compared to previous studies such as Ramondo and Rodríguez-Clare (2013) or Arkolakis et al. (2018). These differences arise from the fact that we allow for *fixed* costs in addition to MP cost. We use affiliate count data to separate fixed costs from MP cost, while previous studies that only use MP sales and trade data cannot separate these two costs. In the data, there are few MP affiliates, but they account for a large sales volume in their host countries; to match these patterns, we estimate large fixed costs and therefore smaller MP costs. In line with this intuition, when we re-estimate our model without fixed costs and target only MP sales but not MP affiliate stocks, we recover the large MP cost estimates from previous studies (cf. Online Appendix OA.8.2). The presence of economically significant fixed costs is in line with Hjort et al. (2022) which finds that, in real terms, labor compensation of middle management is 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 trade costs, MP costs, and the bilateral component of the fixed costs on our set of gravity variables to understand their determinants (cf. Table OA.2). 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 Guillouet

Figure 6: Trade Costs, MP Costs, and the Bilateral Component of Fixed Cost in the Baseline Calibration



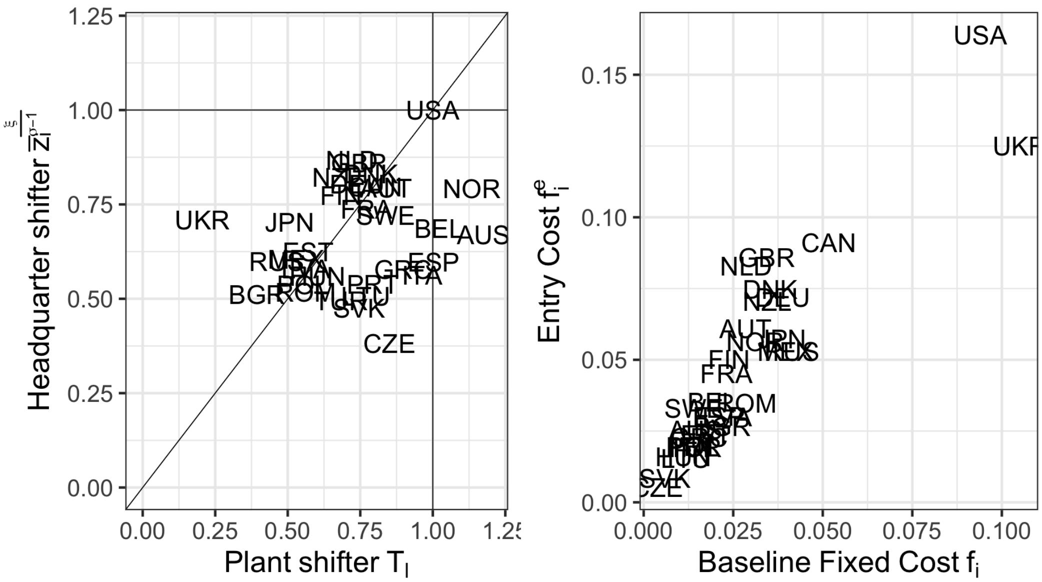
Notes: The figure shows a histogram of the three bilateral cost matrices in the model: trade costs, MP costs, and the bilateral component of fixed costs. We omit the own-country costs which are normalized to 1 for all three types of costs. For MP and the bilateral component of fixed costs, we also omit country pairs where MP is zero, since we set the MP costs to be infinity in those cases.

et al. (2023). Our cost estimates are also consistent with the finding in Alviarez et al. (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, increase little with distance in our baseline model. In contrast, in our alternative calibration of the model without fixed costs, firm market shares respond significantly to distance.

In the left panel of Figure 7, we plot our estimates of the headquarter productivity shifter $\underline{z}_i$ of the firm Pareto distribution against the location productivity shifter T_ℓ of the Fréchet distribution of location-input-specific productivity shocks. Recall that $\underline{z}_i$ is identified by the total amount of foreign affiliate sales of firms headquartered in country i, while T_ℓ is chosen to match countries' level of GDP per capita. Unsurprisingly, the US has the highest headquarter productivity in our dataset. Relatively developed economies with little MNE activity, such as Greece and Portugal, lie below the US and to the right of the 45 degree line that defines the US comparative advantage. On the opposite side of the 45 degree line and close to the US lie developed countries with relatively high MNE activity such as Netherlands, Norway, Finland, and Japan.

In the right panel of Figure 7, we plot the entry cost and the base component of the fixed cost for each country. In our data, one-year survival rates range from 77 to 93 percent, so that the ratio between the base component of the fixed cost and the entry cost varies little across countries.

Figure 7: Technology, Base Component of Fixed Costs, and Entry Costs in the Baseline Calibration



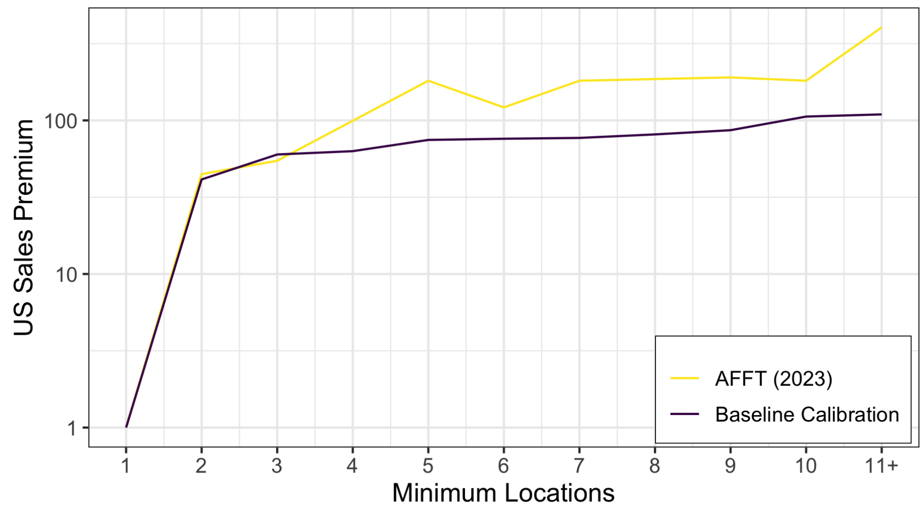
Notes: The figure shows a number of calibrated shifters in the model. The left panel graphs the Pareto minimum $\underline{z}_i^{\xi/(\sigma-1)}$ of the firm productivity distribution against the scale T_ℓ of Fréchet distribution of location-input-specific productivity shocks. The terms $\underline{z}_i^{\xi/(\sigma-1)}$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

More importantly, there is a strong correlation between the level the base component of the fixed cost and GDP of each country. This relationship is the result of an important pattern in our data: the log number of active enterprises increases with the log country population but less than one for one (coefficient of 0.81). To generate this pattern, our model requires fixed costs and entry costs that are increasing with population.[14] This empirical regularity is in contrast with many tractable models of monopolistic competition with firm entry in which the number of entrants is a linear function of the population, such as Melitz (2003), Chaney (2008), or Arkolakis et al. (2018). Because our model does not imply this relationship, our calibration requires a direct measure of entering enterprises M_i, in addition to a measure of local population H_i, in line with Adão et al. (2020).

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 model and the empirical premia documented by Antràs et al. (2022).

[14]In a dynamic setup as in Melitz (2003), our estimates of firm entry costs correspond to the yearly amortized entry cost.

Figure 8: Multinational Sales Premia and Number of Foreign Affiliates in the Data and the Baseline Calibration



Notes: The figure compares size sales premia in the model and in the US data obtained from Antràs et al. (2022) ("AFFT"). The sales premium is measured as the relative sales of US-based multinational firms compared to non-MNEs. .

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.

In the Online Appendix, we provide additional measures of fit for our calibration using a variety of untargeted and targeted moments (cf. Table OA.1).

5.4 Computational Performance of the Algorithm

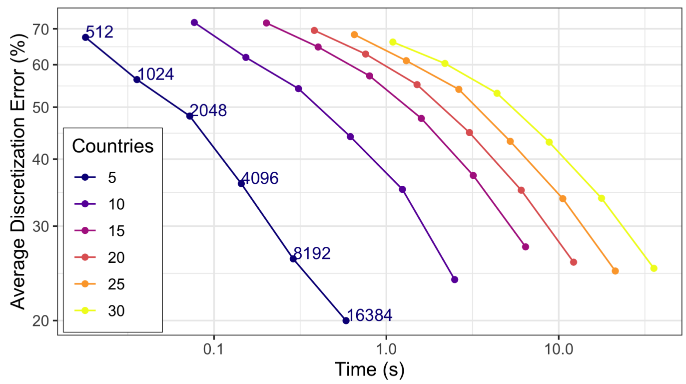
We use the calibrated model to showcase the computational performance of our algorithm using a set of time trials. For trials that vary the number of countries n, we always keep the n largest countries in terms of GDP from our calibrated model, and do not re-calibrate the remaining model parameters. For all trials, we hold general equilibrium objects constant at their calibrated values and document the time it takes to solve for the policy function that maps firm types to optimal location decisions. We conduct three different speed tests, each highlighting different aspects of the algorithm.

In our first set of trials, we compare how long it takes to solve for the full policy function using three different methods: discretize firm productivity with 16384 grid points, then solve the CDCP at each grid point by "brute-force" computing the profit associated with every possible combination of production locations ("Naive"); use the same grid to apply our squeezing and

Figure 9: Comparing Different Methods of Computing the Firm Policy Function

	16384 gridpoints		
	(1)	(2)	(3)
Countries	Naive	Single	Policy
5	1.741	0.584	1.772
10	130.982	2.491	0.699
15	7524.095	6.421	1.558
20	~days	12.198	5.319
25	~mnths	21.293	16.168
30	~yrs	35.657	43.974

Performance by Method in Seconds



Precision-Time Tradeoff

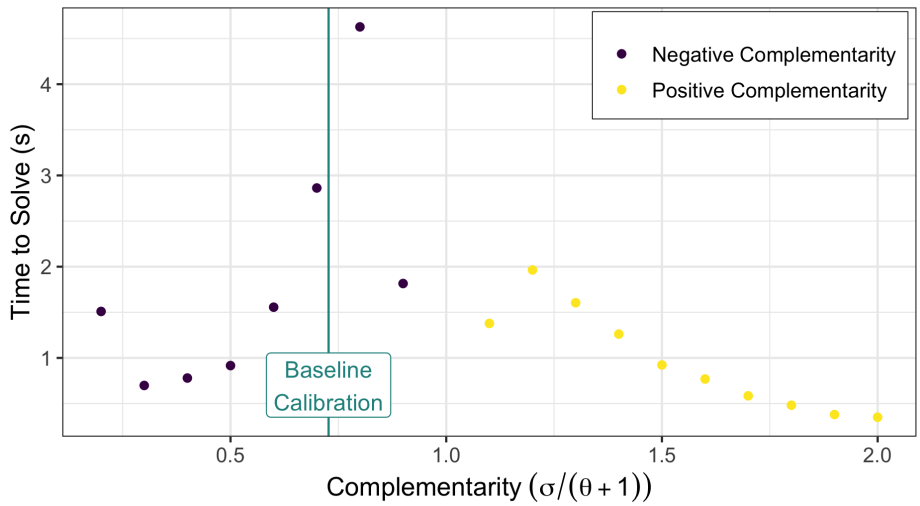
Notes: The table on the left illustrates the computational time for computing the firm policy function with a grid of 16384 points with the brute force method (evaluating the profit function for all possible production location combinations, "Naive"), the method of Theorem 1 (squeezing and branching, "Single"), and the method of Theorem 2 (generalized squeezing and branching, "Policy"). For the n country case, we select the n largest countries in our set by GDP and discard the rest. We use all parameters, fundamentals, and equilibrium aggregates from the baseline calibration of the model. We then compute the time needed to solve for the policy functions with each of the methods. We repeat this process for each row of the table. The figure on the right illustrates the discretization error in computing the policy function with the "single" 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 error ϵ as follows: $\epsilon = (1/N^3)\sum_{i,\ell,n} \mid \frac{\hat{X}_{i\ell n}}{X_{i\ell n}} - 1 \mid$, 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 (i.e, the exactly solved model) and $\hat{X}_{i\ell n}$ is the same object in the model computed using the "Single" method.

branching algorithm above (cf. Theorem 1) to solve the CDCP at each grid point ("Single"); use our generalized squeezing and branching methods (cf. Theorem 2) to solve for the full policy function without grid discretization ("Policy").

The table in Figure 9 presents the algorithm's run time in seconds for each of the three methods and different numbers of countries. With 15 countries, the policy function solution is more than three orders of magnitude faster than the naive approach. As we double the number of countries, the required computation time for the single and naive methods increases at a polynomial, not exponential, rate. The policy function method is slower than the "single" method for larger numbers of countries, because the branching step becomes complex more quickly for the policy function than it does for the single agent method. However, the policy function method is likely preferable in many settings since it provides an *exact* solution of the policy function whereas the "single" method provides a discrete approximation.

In our second set of trials, we study the tradeoff between computational speed and accuracy. Our measure of the "Single" method's discretization error is the average percentage deviation across all trilateral flows relative to the flows generated by computing the same model using

Figure 10: The Strength and Direction of Complementarities and Computational Performance



Notes: The figure shows the computational time to solve the firm policy function with different degrees of complementarity using the generalized branching method of Theorem 2 (generalized squeezing and branching, "Policy"). The strength of complementarity is measured by the ratio $\sigma/\ (\theta + 1)$, and varied by holding $\sigma = 4$ fixed (its value in the baseline calibration) and varying θ. The baseline calibration uses $\theta = 4.5$ and the figure indicates the resulting level of complementarity.

the exact "Policy" technique. The right panel in Figure 9 graphs the computational time of the algorithm against the discretization error for different grid sizes and different total numbers of countries. For example, with 512 grid points and five countries, computation time is low, but the average discretization error is nearly 70%. The precision of the "Single" method is no better than 20% with 16 thousand grid points; compared to a grid of eight thousand points, discretization error improves by almost ten percentage points but at the cost of almost double the computational time. The cost of increasing precision is also polynomial. Furthermore, the time-error frontier shifts rightward at roughly a constant rate in log-log space as the number of countries increases, reflecting the polynomial time of our solution method.

Finally, Figure 10 shows how computational time depends on the strength and direction of complementarities. We solve the model using the "Policy" method above but vary the value of θ to change the ratio of parameters $\sigma/\ (\theta + 1)$ that determines the direction and strength of the complementarities among production locations in our model (see equation 8). Given our calibrated model, computation time is fastest for extreme values of complementarities, either positive or negative. Computation takes longest for moderately negative complementarities. Intuitively, our algorithm exploits the complementarities among production locations to discard certain production location sets without checking them. With stronger complementarities, more potential production location sets can be discarded without evaluating their associated profits than with moderate complementarities. Nevertheless, the model is generally harder to compute with negative complementarities. In Online Appendix OA.7, we provide some intuition for this finding by showing that, with negative complementarities, the set of fixed points of squeezing step is generically larger and hence the (slower) branching step needs to do more work.

6 Counterfactual Exercises

In this final section, we use the model to simulate two recent major political events with economic consequences specifically relating to the operations of multinational firms. The first is the exit of Great Britain from the EU, commonly referred to as Brexit, and the second the sanctions on Russia following the Russian invasion of Ukraine. In the final part of the section, we focus on the role of fixed costs both in the counterfactual simulations and for the evaluation of the gains from openness more generally.

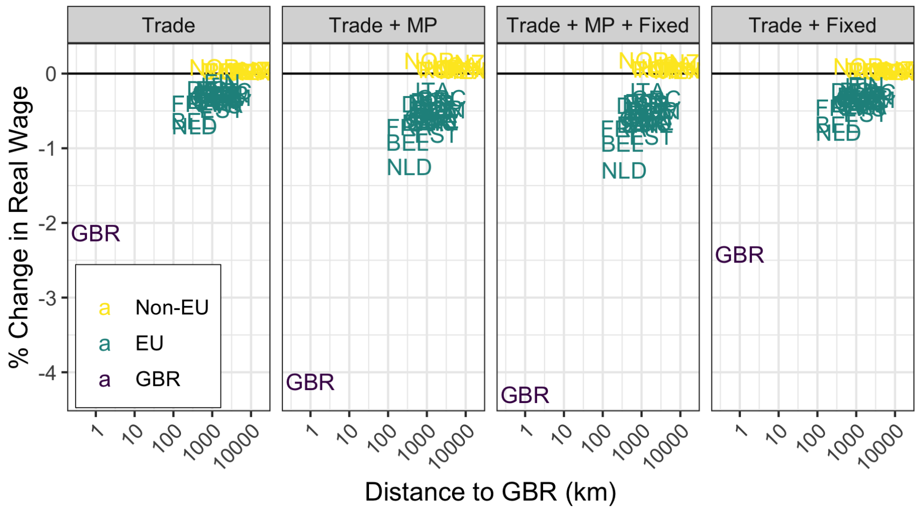
6.1 Great Britain Exits the European Union

We consider three scenarios that capture different kind of barriers to commerce and investment between Great Britain and EU countries: trade barriers, captured by trade costs; technological barriers such as those related to repeal of EU common regulatory environment, captured by MP costs and fixed costs. In the counterfactual exercise below, we increase each of these frictions between Great Britain and EU member countries by 10% one after the other to isolate their individual effects.

Figure 11 shows the impact of these changes on welfare as captured by real wages as a function of a country's distance from Great Britain. European Union member states appear in green, and non-member states in yellow. The first panel shows the effect of raising the barriers to trade by 10%. In this scenario, welfare changes are modest and in line with losses from reductions in trade costs reported in the literature. The most significant losses occur in Great Britain, where the real wage drops by 2%. At the same time, the magnitude of losses of EU countries that face increased trade costs with Great Britain depends on their trading relationship with Britain, of which, given our estimates, their bilateral distance is the most important determinant (see Table A.3 in the Appendix). Great Britain's closest trading allies, such as Belgium, France, and the Netherlands, suffer the most. Countries outside of the EU are barely affected or even have marginal positive gains.

Once variable MP costs also rise, the impact on welfare grows. As we explain below, in reaction to the increased MP costs between Great Britain and EU members, fewer firms based in the EU operate production locations in Great Britain. The resulting lower labor demand in Great Britain exerts downward pressure on wages. At the same time, since fewer EU firms produce in Great Britain, their goods can reach the country only via trade, putting upward pressure on the price index in Great Britain. A similar effect occurs in EU member states as firms based in Great

Figure 11: Real Wages Changes Across Countries in the Brexit Simulation



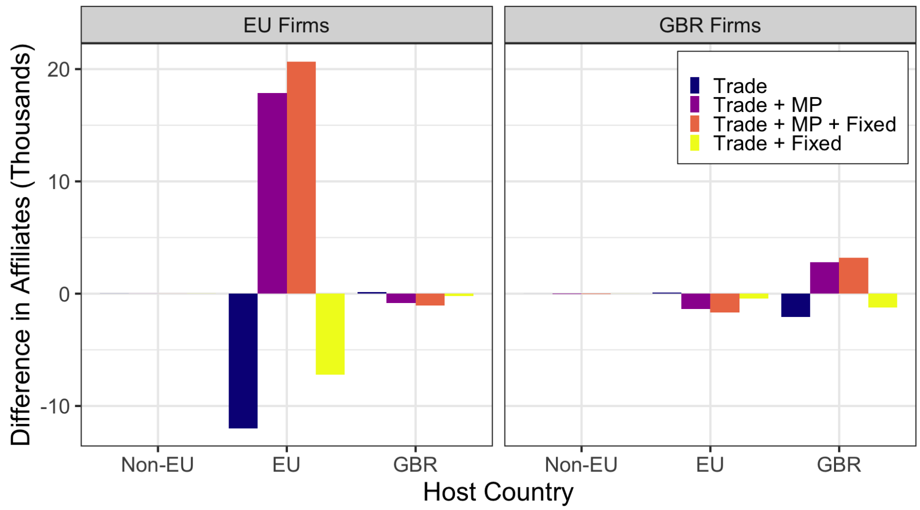
Notes: The figures shows the change in real wage in our Brexit simulation. The effect on EU member countries is shown in green, the effect on non-EU countries is shown in yellow, and the effect on Great Britain is shown in purple. In the first panel, only trade costs increase by 10%, the second panel adds a 10% increase of the MP costs, and the third adds a 10% increase in fixed costs. The fourth panel considers the trade and fixed cost increases in isolation.

Britain withdraw production from EU member states. Effects of geography are again evident as the consequences are more substantial in countries near Great Britain. Adding on increases to the bilateral component of the fixed costs between Great Britain and EU member states in the third panel accentuates this pattern.

To gain more insight into these shifts in production patterns, we plot the mass of operating affiliate sites in Figure 12. The figure displays the change in the mass of affiliates opened by firms headquartered in an EU member state, in host countries that are non-members, members, or Great Britain. The increase in trade costs leads to a small increase in the mass of EU-operated affiliates in Great Britain, accompanied by a decrease in the mass of EU-operated affiliates in the EU and (negligible) decreases in non-EU member states. The increase of British affiliates in the EU is commensurate. This effect reflects a proximity-concentration style trade-off, in line with the evidence in Brainard (1997) and Antràs and Yeaple (2014). Here, negative complementarities play a critical role: as production sites are substitutes, the increase in trade costs leads to EU firms magnifying their in-country presence in Britain and vice-versa. Interestingly, such an effect is not present in a version of our model with positive complementarities where instead, EU-operated affiliates decrease their presence in Great Britain and British firms in the EU (cf. Figure OA.9 in the Online Appendix).

Increases in MP costs and fixed costs have instead a negative impact on the number of EU affiliates in Great Britain. There is again a similar effect for British firms. In addition, the rise of

Figure 12: Net Changes in Affiliates Counts by Sender and Host Country in the Brexit Simulation



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

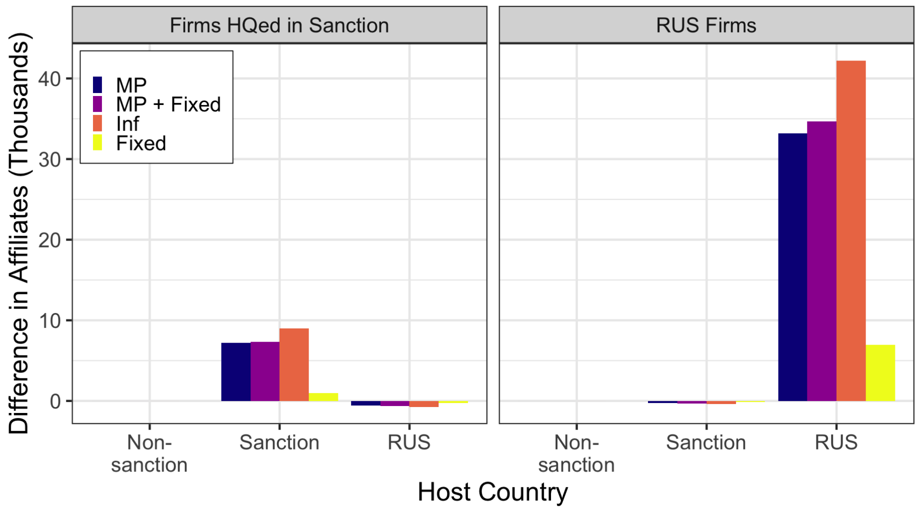
MP costs, both iceberg and fixed, creates a substantial home market effect that leads EU firms to bring production sites back into EU territory. The home-market effect more than compensates for the decrease in the number of foreign affiliates due to the increase in trade costs. However, the most productive foreign affiliates give way to relatively less efficient domestic firms in the EU and Great Britain. The combination of negative nominal wage responses and a significant increase in Great Britain's consumer price index generates significant welfare losses.

6.2 The Sanctions on Russia

We simulate a counterfactual that resembles the sanctions placed on Russia and the in-kind retaliation of the Russian Federation. In our country set, the sanctioning countries are the USA, Canada, the European Union countries, and Australia. In particular, we progressively impose a 30% increase in the MP cost, an additional 30% increase in the fixed cost, and finally, the presently not-so-unrealistic scenario of infinite MP costs, which altogether prohibit MP between the sanctioning countries and Russia.

To begin, Figure 13 summarizes the effect of the policy on affiliate counts. In Figure OA.2 of the Online Appendix, we plot these effects as percentage changes. The panel on the left describes the changes in the number of affiliates operated by firms headquartered in countries that placed sanctions. As the sanctions are progressively introduced, these firms withdraw production from

Figure 13: Net Changes in Affiliate Counts by Sender and Host Country in the Sanctions on Russia Simulation



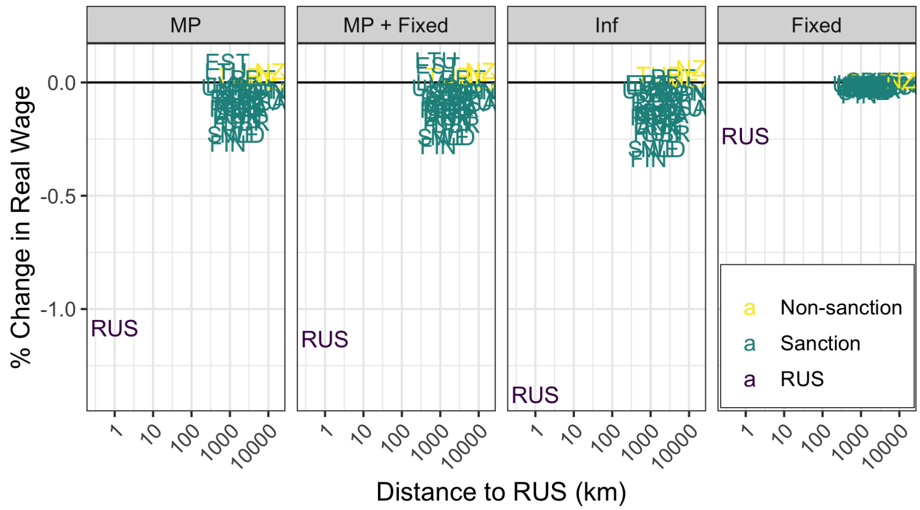
Notes: The left panel of the figure shows the change in the mass of affiliates operated by firms headquartered in countries that placed sanctions to Russia in Russia itself, within the territory of the sanctioning countries, and in third party countries. The right panel of the figure shows the change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

Russia. The increases in MP costs alone imply that the number of affiliates in Russia operated by firms based in countries placing sanctions drops by more than 70%. In absolute numbers, the presence of multinationals in Russia was small, and thus, the actual number of firms from sanctioning countries that depart Russia is small.[15] Increasing the fixed costs also modestly amplifies this effect. As multinational productions costs continue to rise to infinity, all previous affiliates operated by these firms in Russia must close, so that the mass of affiliates in Russia drops to zero. By comparison, there is little change in their production presence among other countries.

The panel on the right describes the production presence of firms headquartered in Russia. The main effect of the sanctions is to decrease the number of affiliates operated by Russian firms in countries placing sanctions. In tandem, the mass of Russian-owned affiliates increases in both countries not implementing sanctions and Russia itself. The change in affiliate counts in countries not placing sanctions is modest in levels, but represents a more than 25% increase over the baseline, as shown Figure OA.2. This adjustment is in line with the export platform motive of FDI. Furthermore, the rise in production sites operated domestically reflects the loosening of competition in both the Russian labor market and the Russian goods market, triggering negative

[15] According to https://www.yalerussianbusinessretreat.com/, 71% of US multinational firms operating in Russia have already withdrawn or suspended their operations as of June 1st, 2023.

Figure 14: Real Wages Changes Across Countries in the Sanctions on Russia Simulation



Notes: The figures shows the change in real wage in our Sanctions on Russia simulation. The effect on countries imposing sanctions on Russia is shown in green, the effect on non-sanctioning countries is shown in yellow, and the effect on Russia is shown in purple. In the first panel, only MP costs increase by 30% while the second panel adds a 30% increase in the fixed costs. In the third panel, we set MP costs to infinity. The fourth panel shows the case in which only fixed costs increase by 30%relative to the baseline.

welfare effects. Similarly to the Brexit counterfactual, as a few large and productive firms based in sanctioning countries cease production in Russia, they are replaced by a marge larger number of small-scale domestic Russian enterprises.

Figure 14 shows the effects of these changes in production location decisions for real wages. Russia suffers modest losses since inward and outward MP are a relatively small fraction of Russian GDP. However, countries that impose sanctions and have close geographic proximity to Russia, like the Netherlands and Finland, have non-negligible losses of around 0.25%. MP in Russia uniformly declines by more than 60%-70% for the countries imposing sanctions in our simulations. Countries that do not impose sanctions are affected little as they are geographically remote to Europe but benefit from small increases in the Russian FDI due to diversion effects. Again, complementarities are critical. As we show in Figure OA.10, in the case of positive complementarities, countries like the Netherlands and Estonia are particularly negatively affected. The relatively large impact in these countries reflect their geographic proximity to many other countries that also impose sanctions, and the ensuing geographic network effects as the European Union collectively withdraws production from Russia.

6.3 The Role of Fixed Costs

In this section, we focus on the role of fixed costs in the simulations above. To this end, we carry out the same simulations in a version of the model that is calibrated without fixed costs.

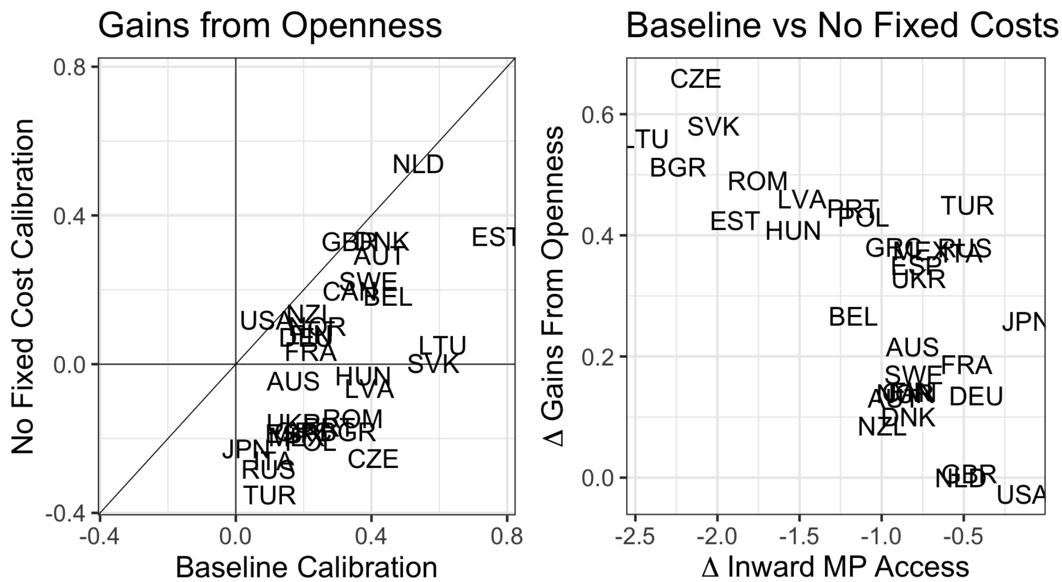
We present this alternative calibration and all associated figures in the Online Appendix but summarize the main findings here. To further understand the role of fixed costs, we also present a comparison of the gains from openness in the baseline calibration of the model and the calibration of the model without fixed costs.

Brexit, Sanctions on Russia, and Fixed Costs The fixed cost increases have small but economically significant effects both in the Brexit and the Russia Sanctions counterfactuals (cf. Figures 11 and 14). The increase in fixed costs tends to trigger the closure of smaller foreign affiliates which nevertheless have the substantial sales needed to cover the fixed costs of the foreign affiliate in the first place.

The general equilibrium implications of the presence of location-specific fixed costs are more nuanced: small changes in trade and MP costs, such as the 10% changes in the Brexit simulations, imply similar changes for the wage, prices, and real wages in the baseline model with fixed costs compared to the alternative calibration without fixed costs. This prediction conforms with standard economic intuition and resonates with a large literature in macroeconomics which argues that, under certain conditions, calibrated general equilibrium models of investment with fixed adjustments costs behave similar to neoclassical investment models (see e.g. Thomas (2002); Veracierto (2002); House (2014)). In our case, however, this similarity is driven by our calibration and our model structure: since we calibrate both the model with and without fixed costs to target the same trade and MP shares from the data, and set the same key elasticities as in Table 2, both calibrated models also turn out to predict similar changes in the (trilateral) trade flows from the Brexit counterfactual. In fact, changes to trilateral flows between the fixed and no fixed cost model generate a correlation of 0.97. In light of the findings in Arkolakis et al. (2012) and Allen et al. (2020), it is then not surprising that the wage, price, and real wages effects of small changes to bilateral costs are similar across these two calibrations of the model. However, large changes in trade and MP costs lead to differing predictions for changes in trilateral sales in the two different calibrations. In the Sanctions on Russia counterfactual, where MP costs are raised by 30%, the correlation of the trilateral flows produced by the model with and without fixed costs drops to 0.87. In this counterfactual, production reallocation decisions are appreciably different between the two calibrations.

Gains from Openness and Fixed Costs We consider the differences between the baseline and the no-fixed cost model in the case of large changes in trade and MP costs by comparing the gains from openness in the two versions of our model. We define gains from openness as the percentage change in the real wage between the calibrated version of the model and the equilibrium of the model with trade and MP autarky, that is $\tau_{\ell n} = \infty$ for all $\ell \neq n$ and $\gamma_{i\ell} = \infty$ for all $i \neq \ell$. The left panel of Figure 15 graphs the gains from openness in the baseline calibration against the gains in the calibrated model without fixed costs. In the calibration without fixed

Figure 15: Fixed Costs and the Gains from Openness



Notes: The left panel shows the gains from openness in our baseline calibrated model and in the alternative calibration without fixed costs. We define gains from openness as the percentage change in the real wage from the calibrated version of the model to the model with trade and MP autarky, that is, $\tau_{\ell n} = \infty \forall \ell \neq n$ and $\gamma_{i\ell} = \infty \forall i \neq \ell$. To construct the y-axis of the right panel, we subtract the gains from openness in the calibration without fixed costs from the gains from openness in the baseline calibration. To construct the x-axis, we first construct a measure of inward MP access. For each production location ℓ, we compute the average inward MP cost $\gamma_{i\ell}$ across headquarter locations i, weighted by the fraction of all sales in location ℓ that are due to affiliates from headquarter location i. The x-axis shows the inward MP access measure in the baseline calibration subtracted from the inward MP access measure in no fixed cost calibration.

costs, the gains from openness are negative for some countries and, even if positive, always strictly lower than those in the baseline calibration of the model.

The different gains from openness arise because, as previously discussed, the calibration without fixed costs requires high MP costs to match the aggregate MP data, while the baseline calibration matches the same data via low MP costs and high fixed costs. Small, less productive countries are more affected by the difference in MP costs in the non-autarky equilibria since they are relatively more reliant on foreign activity, which is inhibited by the high MP costs of the calibration without fixed costs.

To draw out this intuition, we calculate a measure of inward MP market access for each location ℓ: the weighted average MP cost $\gamma_{i\ell}$ across headquarter locations i, where the weights are the fraction of the total value of production in location ℓ accounted for by affiliates of firms headquartered in location i. The right panel of Figure 15 graphs the differences in the gains from openness between the two calibrations against the differences in inward MP market access. It shows that for large, rich countries like the US, there is almost no difference in inward MP market access across the two calibrations because most of the production in the US economy is carried out by firms also headquartered in the US. As a result, the measure of inward MP cost is close to 1 in both calibrations since we normalized the diagonal of the MP cost matrix to 1. In smaller countries such as Slovakia, foreign affiliates from other countries account for a sizable fraction of total production, shifting more weight to the off-diagonal entries of the MP cost matrix, which differ substantially between the baseline and no fixed cost calibrations. These findings are in line with micro evidence in Alviarez et al. (2023), which shows that the local market shares of MNEs are systematically larger in less-developed countries. Taken together, the gains from openness are dramatically lower for small open countries in the calibration without fixed costs. Models of MNE activity that omit fixed costs may systematically underestimate the gains from openness in smaller, less productive countries that depend on foreign multinational production.

7 Conclusion

We introduce a new methodology to solve combinatorial discrete choice problems with negative or positive complementarities and aggregate optimal choices across heterogeneous agents. We put our new methods to work to solve and calibrate a quantitative model of multinational location and production decisions. We conduct two counterfactual simulations. The first resembles the exit of Great Britain from the EU, and we find that, in response to increased trade, MP, and fixed costs, many British firms shut down their operations in EU countries, and EU firms moved out of Britain, generating large losses from Brexit for Great Britain and many European countries. The second counterfactual simulates the recent sanction war between Western countries and Russia in response to the Russian invasion of Ukraine. In response to drastic increases in trade and MP costs, we find that the exit of European firms from Russia plays a central role in understanding the economic implications of the sanctions on the Russian Federation.

References

ADÃO, R., C. ARKOLAKIS, AND S. GANAPATI (2020): "Aggregate Implications of Firm Heterogeneity: A Nonparametric Analysis of Monopolistic Competition Trade Models," Tech. rep., National Bureau of Economic Research.

ALFARO-URENA, A., J. CASTRO-VINCENZI, S. FANELLI, AND E. MORALES (2023): "Firm Export Dynamics in Interdependent Markets," Tech. rep., National Bureau of Economic Research.

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

ALVIAREZ, V. (2019): "Multinational Production and Comparative Advantage," *Journal of International Economics*, 119, 1–54.

ALVIAREZ, V., J. CRAVINO, AND N. RAMONDO (2023): "Firm-Embedded Productivity and Cross-Country Income Differences," *Journal of Political Economy*, 131, 2289–2327.

AMITI, M., O. ITSKHOKI, AND J. KONINGS (2019): "International Shocks, Variable Markups, and Domestic Prices," *The Review of Economic Studies*, 86, 2356–2402.

ANTRÀS, P., E. FADEEV, T. C. FORT, AND F. TINTELNOT (2022): "Global Sourcing and Multinational Activity: A Unified Approach," Tech. rep., National Bureau of Economic Research.

ANTRAS, P., T. C. FORT, AND F. TINTELNOT (2017): "The Margins of Global Sourcing: Theory and Evidence from US Firms," *American Economic Review*, 107, 2514–64.

ANTRÀS, P. AND S. R. YEAPLE (2014): *Multinational Firms and the Structure of International Trade*, vol. 4, 55–130, from January 2013.

ARKOLAKIS, C. (2010): "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 118, 1151–1199.

——— (2016): "A Unified Theory of Firm Selection and Growth," *The Quarterly Journal of Economics*, 131, 89–155.

ARKOLAKIS, C., A. COSTINOT, D. DONALDSON, AND A. RODR ÍGUEZ-CLARE (2019): "The Elusive Pro-Competitive Effects of Trade," *The Review of Economic Studies*, 86, 46–80.

ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

ARKOLAKIS, C., N. RAMONDO, A. RODRÍGUEZ-CLARE, AND S. YEAPLE (2018): "Innovation and Production in the Global Economy," *American Economic Review*, 108, 2128–73.

BALINSKI, M. L. (1965): "Integer Programming: Methods, Uses, Computations," *Management Science*, 12, 253–313.

BEHRENS, K., G. MION, Y. MURATA, AND J. SUEDEKUM (2020): "Quantifying the Gap Between Equilibrium and Optimum under Monopolistic Competition," *The Quarterly Journal of Economics*, 135, 2299–2360.

BERNARD, A. B., J. B. JENSEN, S. J. REDDING, AND P. K. SCHOTT (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 21, 105–130.

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

BOEHM, C. E., A. FLAAEN, AND N. PANDALAI-NAYAR (2023): "New Measurement of Export Participation in US Manufacturing," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 113, 93–98.

BRAINARD, S. L. (1997): "An Empirical Assessment of the Proximity-Concentration Trade-off Between Multinational Sales and Trade," *The American Economic Review*, 87, 520–544.

BRODA, C. AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 121, 541–585.

CHANEY, T. (2008): "Distorted Gravity: the Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–1721.

CONTE, M., P. COTTERLAZ, T. MAYER, ET AL. (2023): *The CEPII Gravity Database*.

CORNUÉJOLS, G., G. NEMHAUSER, AND L. WOLSEY (1983): "The Uncapicitated Facility Location Problem," Tech. rep., Cornell University Operations Research and Industrial Engineering.

COSTINOT, A. (2009): "An Elementary Theory of Comparative Advantage," *Econometrica*, 77, 1165–1192.

DE LOECKER, J., P. K. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNIK (2016): "Prices, Markups, and Trade Reform," *Econometrica*, 84, 445–510.

DYRDA, S., G. HONG, AND J. B. STEINBERG (2023): "A Macroeconomic Perspective on Taxing Multinational Enterprises," Tech. rep.

EATON, J. AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

EATON, J., S. KORTUM, AND F. KRAMARZ (2011): "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 79, 1453–1498.

FAJGELBAUM, P. D., E. MORALES, J. C. SUÁREZ SERRATO, AND O. ZIDAR (2019): "State Taxes and Spatial Misallocation," *The Review of Economic Studies*, 86, 333–376.

FAN, Y. AND C. YANG (2020): "Competition, Product Proliferation, and Welfare: A Study of the US Smartphone Market," *American Economic Journal: Microeconomics*, 12, 99–134.

FEENSTRA, R. C. (2003): "A Homothetic Utility Function for Monopolistic Competition Models, Without Constant Price Elasticity," *Economics Letters*, 78, 79–86.

——— (2018): "Restoring the Product Variety and Pro-Competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity," *Journal of International Economics*, 110, 16–27.

FEENSTRA, R. C., R. INKLAAR, AND M. P. TIMMER (2015): "The Next Generation of the Penn World Table," *American Economic Review*, 105, 3150–3182.

GOLDBERG, P. K. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the US Automobile Industry," *Econometrica*, 891–951.

GUADAGNI, P. M. AND J. D. LITTLE (1983): "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2, 203–238.

GUILLOUET, L., A. KHANDELWAL, R. MACCHIAVELLO, AND M. TEACHOUT (2023): "Language Barriers in Multinationals and Knowledge Transfers," Tech. rep., National Bureau of Economic Research.

HEAD, K. AND T. MAYER (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook," in *Handbook of International Economics*, Elsevier, vol. 4, 131–195.

——— (2019): "Brands in Motion: How Frictions Shape Multinational Production," *American Economic Review*, 109, 3073–3124.

HELPMAN, E., M. J. MELITZ, AND S. R. YEAPLE (2004): "Export versus FDI with Heterogeneous Firms," *American economic review*, 94, 300–316.

HENDEL, I. (1999): "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns," *The Review of Economic Studies*, 66, 423–446.

HJORT, J., H. MALMBERG, AND T. SCHOELLMAN (2022): "The Missing Middle Managers: Labor Costs, Firm Structure, and Development," Tech. rep., National Bureau of Economic Research.

HOLMES, T. J. (2011): "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 79, 253–302.

HOUSE, C. L. (2014): "Fixed Costs and Long-Lived Investments," *Journal of Monetary Economics*, 68, 86–100.

JIA, P. (2008): "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76, 1263–1316.

KIMBALL, M. S. (1995): "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 27, 1241–1277.

KLENOW, P. J. AND J. L. WILLIS (2016): "Real Rigidities and Nominal Price Changes," *Economica*, 83, 443–472.

KLIMEŠ, J. (1981): "Fixed Edge Theorems for Complete Lattices," *Archivum Mathematicum*, 17, 227–234.

KRARUP, J. AND P. M. PRUZAN (1983): "The Simple Plant Location Problem: Survey and Synthesis," *European Journal of Operational Research*, 12, 36–81.

KRUGMAN, P. R. (1979): "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 9, 469–479.

KUSRAEV, A. G. AND S. S. KUTATELADZE (2012): *Boolean Valued Analysis*, vol. 494, Springer Science & Business Media.

LIND, N. AND N. RAMONDO (2023): "Trade with Correlation," *American Economic Review*, 113, 317–53.

MATSUYAMA, K. AND P. USHCHEV (2017): "Beyond CES: Three Alternative Classes of Flexible Homothetic Demand Systems," *Global Poverty Research Lab Working Paper*.

MCFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

MILGROM, P. R. (2004): *Putting Auction Theory to Work*, Cambridge University Press.

MORALES, E., G. SHEU, AND A. ZAHLER (2019): "Extended Gravity," *The Review of Economic Studies*, 86, 2668–2712.

OBERFIELD, E., E. ROSSI-HANSBERG, P.-D. SARTE, AND N. TRACHTER (2023): "Plants in Space," Tech. rep., National Bureau of Economic Research.

OWEN, S. H. AND M. S. DASKIN (1998): "Strategic Facility Location: A Review," *European Journal of Operational Research*, 111, 423–447.

POLLAK, R. A. (1971): "Additive Utility Functions and Linear Engel Curves," *The Review of Economic Studies*, 38, 401–414.

RAMONDO, N. (2014): "A Quantitative Approach to Multinational Production," *Journal of International Economics*, 93, 108–122.

RAMONDO, N. AND A. RODRÍGUEZ-CLARE (2013): "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, 121, 273–322.

RAMONDO, N., A. RODRÍGUEZ-CLARE, AND F. TINTELNOT (2015): "Multinational Production: Data and Stylized Facts," *American Economic Review*, 105, 530–536.

SILVA, J. S. AND S. TENREYRO (2006): "The Log of Gravity," *The Review of Economics and statistics*, 88, 641–658.

SIMONOVSKA, I. (2015): "Income Differences and Prices of Tradables: Insights from an Online Retailer," *The Review of Economic Studies*, 82, 1612–1656.

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

TARSKI, A. (1955): "A Lattice-Theoretical Fixpoint Theorem and its Applications," *Pacific Journal of Mathematics*, 5, 285–309.

THOMAS, J. K. (2002): "Is Lumpy Investment Relevant for the Business Cycle?" *Journal of Political Economy*, 110, 508–534.

TINTELNOT, F. (2017): "Global Production with Export Platforms," *The Quarterly Journal of Economics*, 132, 157–209.

TOPKIS, D. M. (1978): "Minimizing a Submodular Function on a Lattice," *Operations research*, 26, 305–321.

TRAIN, K. (1986): *Qualitative Choice Analysis: Theory, Cconometrics, and an Application to Automobile Demand*, vol. 10, MIT press.

VERACIERTO, M. L. (2002): "Plant-Level Irreversible Investment and Equilibrium Business Cycles," *American Economic Review*, 92, 181–197.

VERTER, V. (2011): "Uncapacitated and Capacitated Facility Location Problems," in *Foundations of Location Analysis*, Springer, 25–37.

XIANG, W. (2022): "Clean Growth and Environmental Policies in the Global Economy," Yale University mimeo.

YEAPLE, S. R. (2013): "The Multinational Firm," *Annual Review of Economics*, 5, 193–217.

ZHENG, F. (2016): "Spatial Competition and Preemptive Entry in the Discount Retail Industry," *Columbia Business School Research Paper*.

A Appendix: Proofs and Additional Results

In this section, we provide proofs for Theorems 1 and 2 and introduce Propositions referenced in the body of the paper. We begin with the proof of Theorem 1.

Proof. The proof proceeds by induction to show that successively applying the squeezing step weakly narrows down the choice space without eliminating the optimal decision set.

Suppose the return function satisfies SCD-C from above. Denote the bounding sets after the kth application by $[\underline{\mathcal{L}}^{(k)}, \overline{\mathcal{L}}^{(k)}]$. Starting with $[\emptyset, L]$, it is trivially the case that $\emptyset \subseteq \underline{\mathcal{L}}^{(1)}$ and $\overline{\mathcal{L}}^{(1)} \subseteq L$. What remains to show is that the upper bounding set contains $\mathcal{L}^\star$ and that the lower bounding set contains no values that are not in $\mathcal{L}^\star$. Let $\ell \in \underline{\mathcal{L}}^{(1)}$. Then, $D_\ell \pi(L) > 0$ so $D_\ell \pi(\mathcal{L}^\star) > 0$ by SCD-C from above. So $\ell \in \mathcal{L}^\star$. Since ℓ was an arbitrary element of $\underline{\mathcal{L}}^{(1)}$, we conclude $\underline{\mathcal{L}}^{(1)} \subseteq \mathcal{L}^\star$. Next, consider $\ell \in \mathcal{L}^\star$. Then, $D_\ell \pi(\mathcal{L}^\star) > 0$ by optimality, so $D_\ell \pi(\emptyset) > 0$ by SCD-C from above. Thus, $\ell \in \overline{\mathcal{L}}^{(1)}$ and since ℓ was an arbitrary member of $\mathcal{L}^\star$, it must be the case that $\mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(1)}$. We thus establish that $\underline{\mathcal{L}}^{(1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(1)}$.

Now suppose $\underline{\mathcal{L}}^{(k-1)} \subseteq \underline{\mathcal{L}}^{(k)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k)} \subseteq \overline{\mathcal{L}}^{(k-1)}$. We show that $\underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$. Select $\ell \in \underline{\mathcal{L}}^{(k)}$. It must be the case that $D_\ell \pi(\overline{\mathcal{L}}^{(k-1)}) > 0$. Since $\overline{\mathcal{L}}^{(k)} \subseteq \overline{\mathcal{L}}^{(k-1)}$, SCD-C from above implies $D_\ell \pi(\overline{\mathcal{L}}^{(k)}) > 0$, so $\ell \in \underline{\mathcal{L}}^{(k+1)}$. Since ℓ was an arbitrary element of $\underline{\mathcal{L}}^{(k)}$, we conclude $\underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)}$. Similarly, now select $\ell \in \overline{\mathcal{L}}^{(k+1)}$. We show it is in $\overline{\mathcal{L}}^{(k)}$. Because $D_\ell(\underline{\mathcal{L}}^{(k)}) > 0$ and $\underline{\mathcal{L}}^{(k-1)} \subseteq \underline{\mathcal{L}}^{(k)}$, SCD-C from above ensures that that $D_\ell(\underline{\mathcal{L}}^{(k-1)}) > 0$. We conclude $\overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$. We now show $\underline{\mathcal{L}}^{(k+1)}$ and $\overline{\mathcal{L}}^{(k+1)}$ sandwich the optimal decision set. Let $\ell \in \underline{\mathcal{L}}^{(k+1)}$ so that $D_\ell(\overline{\mathcal{L}}^{(k)}) > 0$. By the inductive assumption $\mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k)}$, so SCD-C from above allows us to conclude that $D_\ell(\mathcal{L}^\star) > 0$, implying $\ell \in \mathcal{L}^\star$. Similarly, suppose $\ell \in \mathcal{L}^\star$ so that $D_\ell(\mathcal{L}^\star) > 0$. By the inductive assumption, $\underline{\mathcal{L}}^{(k)} \subseteq \mathcal{L}^\star$, so SCD-C from above implies $D_\ell(\underline{\mathcal{L}}^{(k)}) > 0$, ensuring $\ell \in \overline{\mathcal{L}}^{(k+1)}$.

We thus establish that $\underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$ holds if $\underline{\mathcal{L}}^{(k-1)} \subseteq \underline{\mathcal{L}}^{(k)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(k)} \subseteq \overline{\mathcal{L}}^{(k-1)}$holds. Since we established that $\underline{\mathcal{L}}^{(1)} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(1)}$, by induction, we have proved the claim.

A similar argument follows for the case where SCD-C from below holds.

The squeezing procedure must complete in under $|L|$ iterations. To see this note that at each application the squeezing step has to add at least one additional item to the lower bounding set or exclude an additional item from the upper bounding set. The algorithm never removes items from the lower bounding set or adds items to the upper bounding set. So if we start with the empty set as lower bounding set and the full set of items as the upper bounding set, the squeezing step can be applied at maximum of $|L|$ times before a fixed point is reached. □

Next, we prove Theorem 2. The proof uses the following auxiliary mapping from the main text:

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

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

Proof. Consider the 4-tuple $[(\underline{\mathcal{L}}_0, \overline{\mathcal{L}}_0, M), Z]$ and suppose the return function obeys SCD-C from above and SCD-T. We show that the generalized squeezing step exhaustively partitions Z into disjoint subregions, so that the new 4-tuples induce functions $\underline{\mathcal{L}}(\cdot)$ and $\overline{\mathcal{L}}(\cdot)$ over Z. We then show $\underline{\mathcal{L}}_0 \subseteq \underline{\mathcal{L}}(\mathbf{z}) \subseteq \mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}(\mathbf{z}) \subseteq \overline{\mathcal{L}}_0$ for every $\mathbf{z} \in Z$.

Consider the output of applying the squeezing step in Definition 7. First, observe that $\Lambda_\ell(\overline{\mathcal{L}}_0)$ and $\Lambda^c_\ell(\underline{\mathcal{L}}_0)$ are disjoint. For any type $\mathbf{z}$, receiving positive benefit from ℓ's addition to $\overline{\mathcal{L}}$ implies receiving positive benefit from ℓ's addition to $\underline{\mathcal{L}}$ given SCD-C from above. $\Lambda_\ell(\underline{\mathcal{L}}_0) \setminus \Lambda_\ell(\overline{\mathcal{L}}_0)$ is clearly disjoint from the other two output sets, since it is the complement of the latter and explicitly excludes the former. The three new 4-tuples thus induce a valid partitioning on Z, inducing the functions $\underline{\mathcal{L}}(\cdot)$ and $\overline{\mathcal{L}}(\cdot)$.

Next, it is trivially the case that $\underline{\mathcal{L}}_0 \subseteq \underline{\mathcal{L}}(\mathbf{z})$ for all $\mathbf{z} \in Z$ since $\underline{\mathcal{L}}(\mathbf{z})$ is either $\underline{\mathcal{L}}_0$ or $\underline{\mathcal{L}}_0 \cup \{\ell\}$. A similar argument establishes that $\overline{\mathcal{L}}(\mathbf{z}) \subseteq \overline{\mathcal{L}}_0$ for all $\mathbf{z} \in Z$. We now show that $\underline{\mathcal{L}}(\mathbf{z}) \subseteq \mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}(\mathbf{z})$ for all $\mathbf{z} \in Z$. Consider an element in $\underline{\mathcal{L}}(\mathbf{z})$. It is either an element in $\underline{\mathcal{L}}_0$, which is a subset of $\mathcal{L}^\star(\mathbf{z})$ by assumption, or it is ℓ. In particular, ℓ is in $\underline{\mathcal{L}}(\mathbf{z})$ only for $\mathbf{z} \in \Lambda_\ell(\overline{\mathcal{L}}_0)$. These are the types in $\mathbf{z}$ deriving positive marginal value of ℓ's addition to $\overline{\mathcal{L}}_0$. For these types $\mathbf{z}$, $D_\ell \pi(\mathcal{L}^\star(\mathbf{z}), \mathbf{z}) > 0$ by SCD-C from above, since $\mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}_0$ by assumption. Now, we show that $\mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}(\mathbf{z})$ for all $\mathbf{z} \in Z$. Note, again, that $\mathcal{L}^\star(\mathbf{z}) \subseteq \overline{\mathcal{L}}_0$, by assumption. Further, $\overline{\mathcal{L}}(\mathbf{z}) = \overline{\mathcal{L}}_0$ for all 4-tuples except the second, which is associated with subregion $\Lambda^c_\ell(\underline{\mathcal{L}}_0)$. For types $\mathbf{z}$ in this subregion, $\overline{\mathcal{L}}(\mathbf{z}) = \overline{\mathcal{L}}_0 \setminus \{\ell\}$. What remains to show, then, is that $\ell \notin \mathcal{L}^\star(\mathbf{z})$ for types in this subregion. By definition, $D_\ell \pi(\underline{\mathcal{L}}_0, \mathbf{z}) < 0$ for types in this region, so it must be that $D_\ell(\mathcal{L}^\star(\mathbf{z}), \mathbf{z}) \leq 0$ by SCD-C from above. We therefore conclude that $\ell \notin \mathcal{L}^\star(\mathbf{z})$ for these types.

A similar argument follows for return functions exhibiting SCD-C from below instead of SCD-C from above. □

B Data and Calibration

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

B.1 Data

Trade, Foreign Affiliate Sales, and Foreign Affiliate Counts For all of our bilateral flow data, we use the dataset compiled by Alviarez (2019). The data set combines information from four major databases: (1) OECD International Direct Investment Statistics and the Statistics on Measuring Globalization; (2) Eurostat Foreign Affiliate Statistics database; (3) Bureau of Economic Analysis (BEA) public data; and (4) 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: 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. 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 French companies in Germany; note that there is no information on the destination countries of foreign affiliates in a given country, i.e., we do not know how much French 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 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 et al. (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.

B.2 Calibration

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

B.2.1 Estimating Equation Specification for PPML

Here we show the estimating equation underlying Table 1:,

$$y_{ij}^x = \exp\left(\alpha^x + \beta_d^x \log d_{ij} + \beta_{\text{COL}}^x \text{COL}_{ij} + \beta_{\text{COM}}^x \text{COM}_{ij} + \beta_{\text{BOR}}^x \text{BOR}_{ij} + \delta_x' X_{ij} + \varkappa_i + \zeta_j\right) + \epsilon_{ij}^x, \tag{A.1}$$

where d_{ij} indexes distance between countries i and j, the other reported gravity controls are dummies for colonial relations (COL), common language (COM) and common borders (BOR). The vector X_{ij} contains additional gravity controls that we do not target in our calibration and hence do not reported in the table for conciseness, in particular bilateral tariffs and a free trade agreement dummy. The terms $\varkappa_i$ and ζ_j are origin and destination specific fixed effects. We estimate the equation for three outcome variables indexed by x: trade flows, inward MP and total inward affiliates.

We estimate equation (A.1) via Poisson Pseudo Maximum Likelihood (Silva and Tenreyro (2006)) and present the results in Table 1 in the main part of the paper. Tables A.1 and A.2 show robustness to Table 1 in the main part of the paper. In computing manufacturing tariffs, we had to make a decisions whether to take raw averages of the tariffs of all goods in manufacturing or weight in some way. Table 1 includes unweighted bilateral tariffs as a control. A.1 estimates the same equation but with weighted tariffs where the weights are global trade shares of each good following Arkolakis et al. (2018). Table A.2 shows the results from estimating the specification using OLS instead of Poisson Pseudo Maximum Likelihood, both with unweighted and weighted tariffs.

B.2.2 Parameterizing Bilateral Costs

We choose the following functional forms of the bilateral trade cost, MP cost , and fixed cost:

$$\begin{aligned}
\log \tau_{\ell n} &= \bar{\tau}_n \times \mathbb{1}[\ell \neq n] + \kappa_\tau^d \log d_{\ell n} + \kappa_\tau^{\text{COL}} \text{COL}_{\ell n} + \kappa_\tau^{\text{COM}} \text{COM}_{\ell n} + \kappa_\tau^{\text{BOR}} \text{BOR}_{\ell n} + \log(1 + t_{\ell n}) \\
\log \gamma_{i\ell} &= \bar{\gamma}_\ell \times \mathbb{1}[i \neq \ell] + \kappa_\gamma^d \log d_{i\ell} + \kappa_\gamma^{\text{COL}} \text{COL}_{i\ell} + \kappa_\gamma^{\text{COM}} \text{COM}_{i\ell} + \kappa_\gamma^{\text{BOR}} \text{BOR}_{i\ell} \\
\log \nu_{i\ell} &= \bar{\nu}_\ell \times \mathbb{1}[i \neq \ell] + \kappa_\nu^d \log d_{i\ell} + \kappa_\nu^{\text{COL}} \text{COL}_{i\ell} + \kappa_\nu^{\text{COM}} \text{COM}_{i\ell} + \kappa_\nu^{\text{BOR}} \text{BOR}_{i\ell}
\end{aligned} \tag{A.2}$$

where $d_{\ell n}$ denotes the distance between two countries' most populous cities, $\text{COL}_{\ell n}$ is a dummy for two countries' past or present colonial relation relationship, $\text{COM}_{\ell n}$ is a dummy for common language and $\text{BOR}_{\ell n}$ is a dummy for a shared boarder. Finally, $t_{\ell n}$ is the average manufacturing tariff rate country n imposes on imports from country ℓ. The $\{\bar{\tau}_n, \bar{\gamma}_\ell, \bar{\nu}_\ell\}$ components represent

Table A.1: Trade, MP, and Foreign Affiliate Gravity in the Data with Weighted Tariffs

	Trade (1)	MP (2)	Affiliates (3)
Log Distance	-0.689***	-0.284**	-0.681***
	(0.0540)	(0.107)	(0.0851)
Colony	0.0758	0.000476	0.260
	(0.124)	(0.131)	(0.146)
Contiguity	0.440***	0.412*	0.436***
	(0.0697)	(0.163)	(0.0983)
Language	0.160	0.476**	0.577***
	(0.101)	(0.146)	(0.161)
Observations	992	992	992

Notes: The table presents the estimated coefficients from estimating a standard log linear gravity equation using Poisson Pseudo Maximum Likelihood. The outcome variable differs across the three columns: bilateral manufacturing trade flows (Column 1), bilateral multinational production sales (Column 2), 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. The specifications exclude the diagonal entries of the respective flow matrix. Relative to Table (1) in the paper, this paper uses tariffs that are not raw averages across all manufacturing goods but weighted by the global trade shares of each good. 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.

the costs of doing an activity across borders versus within borders. In our calibration procedure, we estimate the destination-specific components $\{\bar{\tau}_n, \bar{\gamma}_\ell, \bar{\nu}_\ell\}$ by targeting the own-shares of each activity, and the gravity-variable-specific elasticities by targeting the estimated on the corresponding gravity variables in the gravity equations for trade flows, MP, and foreign affiliate stocks. Table A.3 reports the estimated elasticities.

Table A.2: Trade, MP, and Foreign Affiliate Gravity in the Data using OLS

	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

Notes: The table presents the estimated coefficients from estimating a standard log linear gravity equation using OLS. The outcome variable differs across the three columns: bilateral manufacturing trade flows (Column 1), bilateral multinational production sales (Column 2), 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. The specifications exclude the diagonal entries of the respective flow matrix. Relative to Table (1) in the paper, this paper uses OLS instead of PPML to estimate the gravity equation and also shows the case using tariffs that are not raw averages across all manufacturing goods but weighted by the global trade shares of each good. 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.

Table A.3: Estimated Cost Elasticities of the Gravity Variables

	Trade (1)	MP (2)	Affiliates (3)
Log Distance	0.21466	0.00043	0.35448
Colony	-0.02463	0.02816	-0.17593
Contiguity	-0.14191	-0.05796	-0.05731
Language	-0.04597	-0.06507	-0.08173

Notes: The table presents the calibrated elasticities of all gravity variables in each of the bilateral costs in the model as specified in equation (A.2).