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

Costas Arkolakis
Yale University

Fabian Eckert
UCSD

Rowan Shi
Toronto Metropolitan University

First Version: January 2017
This Version: June 2023 [In progress]

We introduce a general quantifiable framework to study the plant location decisions of multinational firms. Firms face fixed costs to set up plants in each potential production location and a general variable elasticity demand function. The firm's plant choice problem is combinatorial because the return on each plant depends on which other plants it also operates. A new computational method to solve such combinatorial discrete choice problems and aggregate optimal solutions across heterogeneous firms is at the core of our analysis. We use a calibrated version of the model to study the exit decision of Britain from the European Union (EU) and recent sanctions on Russia on multinational activity. Multinationals' plant reallocations across countries and geography are central in shaping economic effects in both counterfactuals.

*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 and Yunus Tuncbilek 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, 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.

1. Introduction

In 2016, Multinational enterprises (MNEs) accounted for about a quarter of all global value-added and half of the world's export flows. However, only a tiny fraction of firms are engaged in MNE activity.[1] These facts have made the fixed costs of opening foreign plants a central building block in models of MNE activity (Helpman et al. (2004); Yeaple (2013); Antràs and Yeaple (2014)). In the context of a 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 geographic relationship. To choose the optimal set of plant locations, the firm then has to evaluate all possible combinations of plants separately, which becomes computationally infeasible even with moderate numbers of countries. We refer to such problems as combinatorial discrete choice problems, or CDCP.

This paper provides a general quantitative framework to study plant location decisions that nests most existing MNE treatments in the literature. We introduce a simple iterative procedure to solve the firm's plant location problem in the case of negative complementarities between plants. We also extend the class of problems that can be studied using earlier approaches with positive complementarities between plants. In addition, we provide a method to solve for the policy function that maps firm type to optimal plant strategy in the context of models with many heterogeneous firms that all solve CDCPs. To use the algorithm in our quantitative setup, we provide intuitive sufficient statistics for the firm's problem to exhibit positive and negative between-plant complementarities. We use a calibrated version of our model to study the exit decision of Britain from the European Union (EU). Multinational plant relocation responses imply sizable losses for Britain, some losses for other EU countries, and gains for non-EU members. We also study sanctions on Russia in the aftermath of the invasion of Ukraine and find large reallocations of trade flows, multinational sales, and multinational plants that affect geographically close countries the most.

Firms headquartered in one country pay a fixed cost to set up plants in other countries. Foreign plants are less productive than domestic ones but allow firms to exploit labor-cost differences and save on trade costs to foreign destinations. The firm produces

[1] These numbers are constructed using the OECD AMNE database for a set of countries 36 OECD and 23 non-OECD countries, including the rest of the world.

its final good using a unit continuum of intermediate varieties that all its plants can produce. The firm finds the cheapest supplier among its plants for each destination and intermediate input variety. The substitutability between plants induces a negative "supply-side" complementarity as an individual plant is less likely to be the least-cost supplier for a given destination-variety combination the larger the set of plants the firm operates. On the demand side, a general demand function for the final product of each firm gives rise to a positive "demand-side" complementarity. All else equal, firms with more plants have lower marginal costs and higher sales; each additional plant at lower marginal cost firms is more valuable because the associated marginal cost savings apply to a larger sales volume.

Due to the complementarities between plants, the marginal contribution to the firm's total profits of any plant depends on which other plants the firm operates, making the firm's plant location problem combinatorial. Without further restrictions, such problems quickly become intractable as the space of different combinations of locations (countries) grows exponentially in the number of possible locations. In addition, aggregating the optimal plant decisions across firms of different productivity is challenging since optimal decision sets can vary arbitrarily with firm type. To address the first difficulty, we present a simple iterative procedure to solve combinatorial problems even with many locations as long as the firm's profit function satisfies a "single crossing differences in choices" property or SCD-C. [2] The SCD-C condition restricts the direction of the complementaries between individual plants to (loosely) correspond to either positive or negative complementarities. Our method iteratively eliminates non-optimal decision sets for CDCPs that satisfy the SCD-C condition. Intuitively, our method discards potential plant locations without explicitly computing the profits associated with plant location strategies that include them by evaluating the profit function at points of extreme complementarities.[3]

To address the problem of aggregation, we provide an iterative technique to solve for the set-valued "policy function" which maps firm type into decision sets as long as the firm's profit function satisfies single crossing differences in type, or SCD-T. The SCD-T condition

[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.

[3]For cases where the set of potential solutions is reduced but no single strategy isolated, we define an additional step that often finds the optimal decision set without evaluating the return function for all remaining ones.

restricts how the firm's profit function varies with firm productivity. It guarantees that similar firms have identical optimal decision sets, so the policy function only changes value at discrete points in the type space. Our iterative technique solves for the set-valued "policy function" by identifying these points and the optimal decision sets shared by the types between them. It is then straightforward to integrate the policy function over arbitrary distributions of agent heterogeneity and compute general equilibrium objects.[4]

Our solution and aggregation methods apply to general combinatorial discrete choice problems beyond the particular context of multinational production. In particular, they apply to any maximization problem with an objective function that maps a discrete choice set into a real scalar as long as the objective satisfies the relevant single-crossing conditions.

We return to our quantitative framework with the general solution and aggregation method to show that the firm's maximization problem satisfies SCD-C and SCD-T. We derive explicit conditions under which the problem exhibits positive or negative complementarities. The supply-side channel is governed by a parameter determining the comparative advantage differences among plants. Intuitively, cannibalization across plants is low if comparative advantage differences are large. The strength of the demand-side channel pushing for positive complementarities is governed by the price elasticity of demand and the pass-through elasticity of marginal cost changes to the price whose product determines how marginal cost savings through an additional plant affect the firm's sales. The balance of these two channels determines the type of complementarities exhibited by the firm's objective function.

For our quantitative exercises, we choose a Constant Elasticity of Substitution (CES) demand system which collapses the condition for positive versus negative complementarities to a simple inequality. Our method, which works for both type of complementarities, allows us to estimate the type that best describes the data. We calibrate our model to match trade and multinational production patterns across a set of large countries using a rich data set produced by Alviarez (2019). Fixed costs and discrete decisions imply that our model does not deliver closed-form gravity relationships, different from standard quantitative trade models. We calibrate trade costs, productivity costs of foreign

[4]The policy function can exhibit jumps, non-monotonicities, and partially overlapping decisions sets. Our approach is applicable for continuous and degenerate, single and multidimensional type distributions alike.

production, and the fixed costs of plants to match patterns of trade flows, multinational sales across countries, and the bilateral matrix of foreign plants. In addition, we infer countries' productivities to exactly match differences in GDP per capita across countries.

We first use the calibrated model to conduct a set of speed tests to illustrate the performance of our algorithm. We show how computation time increases in the number of countries and how this depends on the strength and direction of complementarities. We also compare our method to methods that instead solve the firm's problem using brute force, i.e., by evaluating all possible solutions. Overall, our method is drastically faster than existing approaches and enables the solution of negative complementarity problems with many locations that were previously infeasible.

We also conduct a set of counterfactual exercises suggestive of the general equilibrium impacts of Brexit and the European Union (EU)-United States (US) and Russia firm sanctions war. 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 add an increase in variable MP costs by 10%, and finally we incorporate an increase in fixed costs of plant establishment by 10%. When only trade costs increase, nominal wages in Great Britain rise. However, prices increase in tandem, which ultimately results in relatively small effects on real wages. However, once MP frictions are also incorporated, welfare in Great Britain drops appreciably. In reaction to the higher MP costs, EU-based firms withdraw plants from Great Britain, resulting in lower nominal wages and higher prices in Great Britain.

We finally simulate a counterfactual that resembles the sanctions placed on Russia and the retaliation of the Russian Federation. In particular, we simulate a 30% increase in the bilateral iceberg cost of trade and MP between Russia and countries in the EU as well as the USA. We find large reallocations of plants and multinational activity. The effect of geography are salient, as firms from non-sanction countries with headquarters closer to Russia increase their production there by more than firms headquartered in countries far away. Countries that impose sanctions and are close to Russia contract production and shut down plants much more than sanction countries that are far away.

Our paper contributes to a rapidly growing literature in international trade and industrial organization in which firms choose a discrete set of locations to build plants in or source

inputs from.[5] The existing literature offers solutions that either abstract from fixed cost of adding a location and, thus, from CDCPs altogether (see Ramondo (2014), Ramondo and Rodríguez-Clare (2013), Arkolakis et al. (2018), Fajgelbaum et al. (2019)), solves them for a small enough number of discrete locations evaluating all potential plant location strategies by "brute force" (see Tintelnot (2017) and Zheng (2016)), or constrains the analysis in the case of positive complementarities between plants and applied a dimension reduction method introduced by Jia (2008) (see Antras et al. (2017)).[6] Our method opens the way to estimate the direction of complementarities via method of moments which require simulating and solving the model many times and aggregating optimal decision across firms and to conduct general equilibrium counterfactuals once parameters have been estimated.

We also contribute to a literature studying (combinatorial) discrete choice problems by adding a new and unified way of solving single-agent CDCPs with positive or negative complementarities to this literature (cf. Definition 1). In most classical discrete choice problems, agents choose one item from a set of mutually exclusive alternatives (see, e.g., McFadden (1973)). In some applications, agents select multiple items but without complementarities among the individual items and with a pre-specified total number of choices (see, e.g., Hendel (1999)). Jia (2008) is an exception: the paper provides a method to solve combinatorial discrete choice problems in which the objective exhibits supermodularity. Relative to Jia (2008), we present a generalized method applicable to return functions exhibiting a weaker form of positive spillovers — establishing sufficient single crossing conditions instead of the stronger condition of supermodularity – or, more importantly, negative spillovers.[7] A recent paper by Oberfield et al. (2020) solves combinatorial location problems using a continuous limit; their approach is complementary to our discrete locations approach.

A second distinct contribution is our method for solving for the policy function mapping

[5]The classic Simple Plant Location Problem in operations research is an NP-hard problem (see Jakob and Pruzan (1983)). Such problems can be solved with our techniques as they exhibit negative complementarities across the entire choice space.

[6]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)).

[7]As explained above, such negative complementarities are an inherent feature of plant locations problems. See Yang (2020) for a recent application of our method to study the location choices of multi-plant oligopolists in the cement industry, a CDCP with negative complementarities.

heterogeneous agent type to optimal decision, which together with the agent type distribution can be used to aggregate decisions. Aggregation of choices in classical discrete choice models usually rely on agents choosing a single item and a random utility component in the each agents valuation of each item (see Guadagni and Little (1983) and Train (1986)).[8] In the context of multiple discrete choice problems with supermodularity (i.e., CDCPs), existing approaches have relied on discretizing the type heterogeneity and solving the CDCP only for a limited number of types and then interpolating between them (seeAntras et al. (2017)). We show that that such interpolation can lead to large errors in setting with negative complementarities since the policy function exhibits no form of continuity or nesting. We add to this literature by providing a method to solve for the policy function that is exact, without the need for approximation, and usually faster since we directly solve for all "kinks" in the policy function where the optimal decision set changes.

2. A General Framework of Multinational Production

In this section, we present an individual firm's multinational plant location problem taking aggregate demand and the mass of firms headquartered in each country as given. After introducing solution methods for the firm's problem in the next section, we discuss firm entry, market clearing, and the equilibrium system of the model in Section 4.

Setup The world economy consists of a discrete set of countries J. We index firm headquarter locations by i, locations of production by ℓ, and locations of final consumption by n. In each country n, there is a mass L_n of consumers who each inelastically provide 1 unit of labor to firms at a wage w_n. In each headquarter location, there is a mass Ω_i of operating firms. These firms choose whether to establish production plants in other countries ℓ to serve consumers in destinations n, and incur a bilateral fixed cost f_{il}. Firms produce a single differentiated final good, indexed by $\omega \in \Omega \equiv \Omega_1 + \ldots + \Omega_J$, using a unit mass of input varieties produced by their plants, indexed by $\nu \in [0,1]$. Labor markets are perfectly competitive and output markets are monopolistically competitive.

[8]Note that there are random utility type models with correlated shocks across choices, e.g., Bryan and Morten (2019), Arkolakis et al. (2018), and Lind and Ramondo (2018).

Demand System All consumers have identical preferences over the set of final goods. As in Arkolakis et al. (2019), for a given schedule of prices, $\{p_n(\omega)\}_{\omega\in\Omega}$, consumers demand the following quantity of each good:

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

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.[9]

As discussed inArkolakis et al. (2019) the demand expressions in form of equation 1 result from a variety of different utility functions, both homothetic and non-homothetic. In its homothetic formulation, this equation 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, P_n typically represents a price aggregator and Q_n represents aggregate consumption spending. We remain agnostic on restrictions on the aggregator P_n, Q_n for the first part of our analysis. In our quantitative application, we specify an underlying utility function and show the aggregators as explicit functions of prices and demand.

Production Technology Firms produce a single final good that consists of a continuum of firm-specific intermediate varieties combined according to a CES aggregator with an elasticity of substitution denoted by η. Each plant of the firm can potentially produce the entire continuum of intermediate inputs. Firms choose in which countries to operate a plant and then, for each destination market and intermediate input, the least cost supplier among its plants.

The marginal cost of producing an individual input ν in location ℓ for a firm headquartered in location i is given by

$$c_{i\ell}(\nu,\omega) = \frac{\gamma_{i\ell} w_\ell}{\varphi_\ell(\nu,\omega)},$$

[9] In particular, this setup is sufficiently general to include very popular (classes of) demand functions such as the additively separable (Krugman (1979)) and its various parametric specifications, symmetric translog and some generalizations (Feenstra (2003), Feenstra (2018)), as well as Kimball preferences (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).

where $\varphi_\ell(\nu,\omega)$ is a firm-input-plant specific productivity term. The MP cost term γ_{il}, $\gamma_{il} \in (1,\infty)$ with $\gamma_{ii} = 1,\ \forall i$, parameterizes the firm's productivity loss of producing in country ℓ relative to producing in the headquarter location i.

The firm assembles its final composite good separately for each destination market by importing the required intermediate inputs. For each intermediate input, ν, and destination, n, the firm chooses from among its plants the one that offers the lowest destination-specific marginal cost so that:

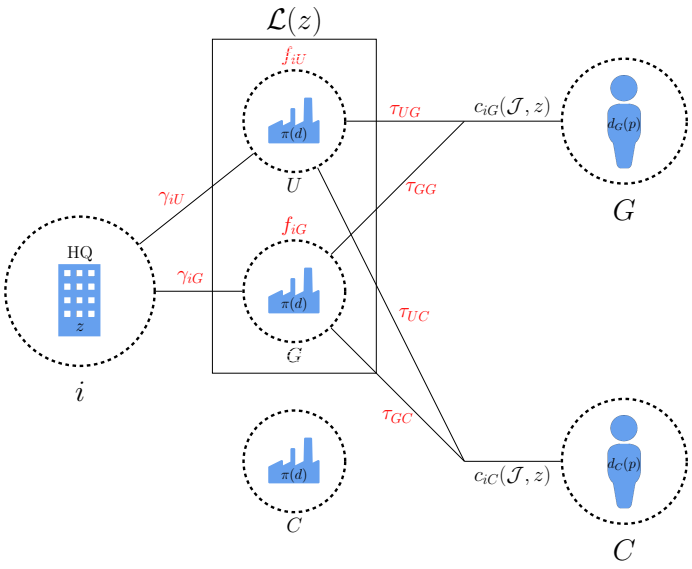
$$c_{in}(\nu;\omega) \equiv \min_{\ell\in\mathcal{L}} c_{i\ell}(\omega)\tau_{\ell n}, \tag{2}$$

is the *realized* unit cost of delivering variety ν to destination n. Here, the term $\tau_{\ell n}$, $\tau_{\ell n} \in (1,\infty)$ with $\tau_{\ell\ell} = 1\,\forall\ell$, is an iceberg cost of shipping one unit of the input from production location ℓ to destination n.

We assume the firm learns about the realization of $\varphi_\ell(\nu,\omega)$ after choosing its plant locations but before choosing the optimal plant ℓ to serve market n. As a result, the firm maximizes its *expected return* when choosing its plant locations $\mathcal{L}$. We follow Tintelnot (2017) in assuming that the firm draws the vector $\boldsymbol{\varphi}_\ell$ for each plant and variety independently from a Fréchet distribution with scale parameter $T_\ell z_\ell(\omega)$ and a shape parameter θ, whose *inverse* indexes the amount of productivity heterogeneity across plants for a given variety and hence serves as a measure of the substitutability of plants in supplying an given variety. The scale parameter includes a country-specific productivity shifter, T_ℓ, common to all firms producing in ℓ and reflects the location's production efficiency. In addition, it contains a firm-specific productivity shifter $z_\ell(\omega)$ indexing the firm's idiosyncratic productivity at producing in location ℓ. We collect a firm's location-specific productivity terms into a vector $\mathbf{z} \equiv \mathbf{z}(\omega) = \{z_1(\omega),\ldots,z_N(\omega)\}$. This distributional assumption yields a closed form expression for the unit cost of a firm of type $\mathbf{z}$ headquartered in location i of delivering a unit of its final good to destination n given its plant network $\mathcal{L}$:

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

FIGURE 1: A Visual Overview 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_{iU} and f_{iG} to operate plants in countries U and G. The firm incurs the efficiency cost of producing in these countries, γ_{iU} and γ_{iG}. The operating plants provide the firm's intermediate input varieties to the output markers in counties G and C subject to the iceberg trade costs τ_{iG} and τ_{iC}, respectively, at average marginal costs of $c_{iG}\left(\mathcal{L};z\right), c_{iC}\left(\mathcal{L};z\right)$, respectively, where $\mathcal{L} = \{G, C\}$.

where $\tilde{\Gamma}$ is a constant of integration.[10] The unit cost increases in wages, MP costs, and trade costs, and decreases in firm productivity and country productivity. We denote $\Theta_{in}(\mathcal{L};\mathbf{z}(\omega))$ as the *production potential* of a particular set of plant locations $\mathcal{L}$ for the destination market n. Figure 1 provides a schematic overview of the model structure. Firms with the same headquarter location and the same productivity $\mathbf{z}$ make identical pricing and plant locations, so for the rest of the paper we index firms by i and $\mathbf{z}$ alone.

[10] If we use instead a plant-specific multivariate Pareto distribution, as in Arkolakis et al. (2018), we obtain a similarly tractable expression. Notice that due to the properties of the Fréchet distribution, the intensive margin does not play a role in final expression, and thus the elasticity of aggregation across varieties, η, is not present in the expression for the cost function, $c_{in}(\mathcal{L};\omega)$.

Profit Maximization The profit maximization problem of the firm has two parts: choosing its price in each destination market given its marginal cost, $\{c_{in}(\mathcal{L};\mathbf{z})\}_n$, and decides on a set of plant locations, $\mathcal{L}$, and uses these plants ("export platforms") to serve final consumers in all countries.

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

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

where ε_D is the elasticity of the demand D and μ is the optimal markup as a function of marginal cost. Equation (4) shows that a firm's markups in market n depends on its full plant location network: plant location choices and markups in each country are co-determined.

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

$$\pi_{in}\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right) \equiv \left(\mu\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right)-1\right) QD\left(p\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right)/P_n\right) c_{in}\left(\mathcal{L};\mathbf{z}\right). \tag{5}$$

The firm chooses $\mathcal{L}$ so as to maximize the sum of all destination-specific profits, so that the firm's total profit is given by:

$$\pi_i(\mathcal{L}^\star;\mathbf{z}) = \max_{\mathcal{L}\in L} \pi_i(\mathcal{L};\mathbf{z}) = \max_{\mathcal{L}\in L}\{\sum_n \pi_{in}\left(c_{in}\left(\mathcal{L};\mathbf{z}\right)\right) - \sum_{\ell\in\mathcal{L}} f_{i\ell}\}. \tag{6}$$

Since there may be complementarities between plants the firm's profit maximization problem is a combinatorial discrete choice problem as in Definition 1.

Complementarities Between Plants There are two distinct sources of complementarity between individual plants in the firms profit function.

First, there is the negative "supply-side" complementarity as apparent in equation (2): the firm's various plants are competing to be the least cost supplier of any given input-destination pair. The more plants the firm operates, the lower the additional sales of any individual plant and hence its marginal value. The strength of this complementarity depends on the substitutability of plants as input suppliers which is measured by the

inverse of Fréchet shape parameter, θ.

Second, there is the positive complementarity which works through the demand system. Note that, all else equal, firms that operate a more extensive set of plants have a lower marginal cost, as reflected by equation (3), and hence larger sales volumes. As a result, the marginal value of each plant for firms with more plants is larger since its associated additional marginal cost savings are applied over a larger sales volume. The strength of this complementarity depends on the elasticity of demand and the pass-through elasticity of costs to price which together determine how changes in marginal cost translate into changes in sales.

In the next section, we introduce a new method to solve the firm's combinatorial discrete choice problem in the presence of positive or negative complementarities that is applicable beyond the context of our particular model as long as a set of conditions on the profit function are met. We also show how to aggregate the discrete choices of heterogeneous firms to solve for market aggregates.

3. Solving and Aggregating CDCPs

We start by defining a general class of combinatorial discrete choice problems which nests the problem solved by the firms in our multinational production framework above.

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

$$\pi(\mathcal{L}; \mathbf{z}) : \mathscr{P}(L) \times \mathbf{Z} \to \mathbb{R}, \tag{7}$$

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

We refer to $\mathcal{L}$ as the firm's "decision set," to L as their "choice set," and to $\mathscr{P}(L)$ as

their "choice space."[11] To formalize the interdependence of decisions, we introduce a marginal value operator that encodes the additional benefit agents derive from element l's inclusion in the decision set $\mathcal{L}$. In Appendix OA.4, we also show that definition (1) nests the canonical Simple Plant Location Problem from the operations research literature.

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

$$D_\ell \pi(\mathcal{L}; \mathbf{z}) \equiv \pi(\mathcal{L} \cup \{\ell\}; \mathbf{z}) - \pi(\mathcal{L} \setminus \{\ell\}; \mathbf{z}). \tag{8}$$

Decisions are *interdependent* as long as the marginal value $D_\ell \pi(\mathcal{L}; \mathbf{z})$ of item l depends on the overall choice set $\mathcal{L}$. If the marginal value of any item ℓ is the same across all decision sets $\mathcal{L}$, decisions are independent and the agent can simply consider each item in isolation. We also define the set-valued policy function that summarizes the optimal decisions of all agents in the economy by mapping an agent's type to their 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 a mapping $\mathcal{L}^\star(\cdot) : \mathbf{Z} \to \mathscr{P}(L)$ such that

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

The policy function is useful for the aggregation of decisions across firms. For any given distribution of firm productivity, $f(z)$, the policy function can be used to compute aggregate quantities and prices in the economy.

We now describe how to solve the CDCP of a *single* agent holding agent type $\mathbf{z}$ fixed. As a first step, we introduce a restriction on the complementarities among choices in the return function called single crossing differences in choices, or SCD-C for short.

[11] The formulation in equation (7) 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 choice of a discrete set $\mathcal{L}$, e.g., conditional on having a plant in location A how much should this plant produce? In such situations, equation 7 describes the CDCP that results after "maximizing out" the continuous choice variable for any given choice set. Likewise equation (7) can be embedded into a dynamic CDCP, where the return to each potential decision set is a value function that takes into accounts its dynamic implications.

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 (7). 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 $\mathcal{L}_1 \subset \mathcal{L}_2 \subseteq L$,

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

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

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

The single crossing restrictions are intuitive. For return functions satisfying SCD-C from above, if the marginal value of including an additional element ℓ in a given decision set is positive, it remains so as elements are removed from the decision set. Similarly, for return functions satisfying SCD-C from below, if the marginal value of including an additional element ℓ in a given decision set is positive, it remains so as other elements are added to the decision set. For the rest of the paper, we will refer to return 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 return functions that exhibit monotone substitutes decreases as more items are added to the decision set. Similarly, for return functions

exhibiting monotone complements, any element's marginal value increases as more items are added to the decision set. In our setting, the definition of monotone substitutes and complements coincides with that of submodularity and supermodularity, respectively.[12]

For return functions satisfying SCD-C, we now present a simple mapping on the choice space associated with a CDCP whose fixed point corresponds to the agent's optimal decision set.

3.2. The Squeezing Procedure

This section presents our "squeezing procedure," a method to solve CDCPs when the 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 eliminates an increasing number of non-optimal decision sets from the choice space and its fixed point always contains the optimal decision set. Since the type vector, $\mathbf{z}$, is held fixed in this section, we omit it for notational brevity.

Consider the choice set L of the CDCP defined in equation (7). 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 items from L that are certain to be included in or excluded from the optimal decision set $\mathcal{L}^\star$. The "subset" $\underline{\mathcal{L}}$ *includes* all items in L we know to *be* in the optimal decision set. The "superset" $\overline{\mathcal{L}}$ *excludes* all items in L 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 items which *may* be in the optimal decision set. We refer to this group as "undetermined" items or elements. A natural starting point for our procedure is to set $\underline{\mathcal{L}} = \varnothing$ and $\overline{\mathcal{L}} = L$, so that $\overline{\mathcal{L}} \setminus \underline{\mathcal{L}} = L$, that is, all items in L are undetermined.

The central mapping 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

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

which collects the items $\ell \in L$ that have a positive marginal value as part of a given

[12] If the choice set is not finite the notions do not coincide. In this case, sub- and supermodularity are implied by monotonicity in set but not vice versa. We provide these results in the Appendix.

decision set $\mathcal{L}$. We then define the squeezing step as follows.

Definition 5. [Squeezing step] Consider a CDCP and its associated bounding sets $[\underline{\mathcal{L}}^{(k)}, \overline{\mathcal{L}}^{(k)}]$.

The mapping S^a is such that

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

the mapping S^b is such that

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

where k indicates the output of the kth application of the squeezing step.

If the underlying return function satisfies SCD-C, each application of the squeezing step adds elements to the subset, $\underline{\mathcal{L}}$, while removing elements from the superset, $\overline{\mathcal{L}}$, thereby eliminating 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 (7).*

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

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

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 applying the squeezing step (weakly) reduces the collection of undetermined items.[13] In particular, the expression $\underline{\mathcal{L}}^{(k)} \subseteq \underline{\mathcal{L}}^{(k+1)}$ implies that (weakly)

[13]The squeezing step is designed to recover $\mathcal{L}^\star$ so that all items ℓ for which the agent is indifferent are excluded from the optimal strategy. If these items should be included, they are easily identified as those ℓ for which $D_\ell \pi(\mathcal{L}^\star) = 0$.

more items are *included* in the subset — and hence known to be in the optimal decision set — after applying the squeezing step. Similarly, the expression $\overline{\mathcal{L}}^{(k+1)} \subseteq \overline{\mathcal{L}}^{(k)}$ implies that (weakly) more items are *excluded* from its superset — and hence known not to be in the optimal decision set — after applying the squeezing step. Crucially, no items that are in the optimal decision 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 operators that indicates applying the mappings S^a and S^b until convergence by $S^{a(K)}$ and $S^{b(K)}$ and by $[\underline{\mathcal{L}}^{(K)}, \overline{\mathcal{L}}^{(K)}]$ the resulting bounding sets. If the converged pair of bounding sets 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 decision set solving the CDCP. [14]

There are cases when the converged pair of bounding sets is not identical, so that there are several potentially optimal solutions. One option is then to apply the computationally expensive brute force approach of manually comparing π across all elements in the remaining choice space. Instead, we introduce a "branching procedure,"described and formally characterized in Appendix OA.2.1, that often finds the optimal decision set much faster than brute force. A convenient property of the branching procedure is that it collapses to the brute force method only in the worst-case scenario.

3.3. Sufficient Conditions for Solving for the Policy Function

In this section, we show how to solve for the policy function mapping agent type into optimal decision set in settings where a large number of heterogeneous agents each solve a CDCP. To that end, we introduce an additional restriction on the return function called single crossing differences in type, or SCD-T for short, that has implications on how the optimal decision set changes with agent type.

We begin by defining the set of all firm efficiency types so that the marginal value of a

[14] 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. When out method is integrated into an estimation routine of the parameters determining the type of SCD-C, it is important to know ex-ante which type of SCD-C a given parameter guess induces in order to choose the appropriate squeezing step.

given location ℓ is positive given a choice set $\mathcal{L}$:

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

We define single crossing differences in type, or SCD-T, a 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 $\Lambda_\ell^c(\mathcal{L})$ are both connected sets.

The two contiguous sets $\Lambda_\ell(\mathcal{L})$ and $\Lambda_\ell^c(\mathcal{L})$ divide the type space $\mathbf{Z}$ into types which receive positive marginal value and types which receive negative marginal value from ℓ's inclusion in $\mathcal{L}$.[15]

Intuitively, the SCD-T restrictions implies that if the addition of item ℓ to choice set $\mathcal{L}$ has positive marginal value for an agent of type $\mathbf{z}$, it also has a positive marginal value for an agent whose type is sufficiently close to $\mathbf{z}$ in the type space. SCD-T restricts agents with similar types to have the same optimal decision set. For illustration, Figure 2 depicts a policy function associated with a one-dimensional type space obeying these restrictions. In the figure, agents with types between z_1 and z_2 have the same optimal decision set $\mathcal{L}_1^\star$, while types between z_2 and z_3 instead optimally choose $\mathcal{L}_2^\star$, and so on. As a result of the SCD-T assumption, the corresponding policy function changes value only at interval boundaries, e.g., z_1 and z_2.

For the case of a one-dimensional type space, we can write the SCD-T restriction in parallel with the SCD-C restriction in Section 3.2. In particular, given two types $z_1 < z_2$, SCD-T asserts, for all elements $\ell \in L$, and decision sets $\mathcal{L}$,

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

When the marginal value function is strictly increasing in the type, z, the return function displays supermodularity between agent type and the decision set (and similarly

[15] The SCD-T property is therefore implied by the supermodularity property introduced by Costinot (2009) (in its log form). Notice that our analysis focuses on partially ordered sets (lattices) not necessarily on totally ordered sets.

submodularity when decreasing).[16]

A sufficient condition for SCD-T follows from super- and sub-modularity between type and decision set. In particular, fix an item ℓ, set $\mathcal{L}$, and component j of the multi-dimensional type vector. Holding all other components of the type vector fixed, 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.[17]

In what follows, we use the SCD-T restriction to solve for the *policy function*, $\mathcal{L}^\star(\cdot)$, that maps agents' types to their optimal decision sets. We re-introduce the $\mathbf{z}$ indexing to indicate an agent's type. Since the value of the return function depends on the agent's type, agents of different types may each have drastically different optimal decision sets.

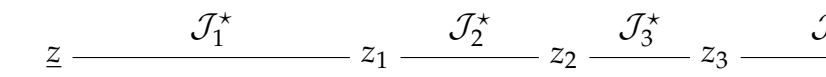
In the special case of single dimensional type heterogeneity with the return function that obeys both SCD-C from below and SCD-T, the policy function obeys a *nesting* structure. That is, given two scalar types $z_1 < z_2$, it must be the case that $\mathcal{L}^\star(z_1) \subseteq \mathcal{L}^\star(z_2)$.[18] In the appendix, we generalize this nesting result to a multidimensional type space under a stronger restriction in place of SCD-T. With SCD-C from *above* instead of *below* the policy function does not necessarily obey a nesting structure: there is no strict "hierarchy of items," with the lowest type agents including only the first, then higher type agents further including the second, and so on. Instead, more productive agents may choose sets that contain less and different items than less productive agents, and vice versa. 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.

[16] Without loss of generality, we assume SCD-T holds from below. If it holds from above, then the problem can be recast in terms of $z' = 1/z$, which exhibits SCD-T from below.

[17] 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.

[18] 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 (Antras et al. (2017) introduced this result to economics). We establish the nesting result in a setting with a multidimensional type space in the case of positive spillovers, but also show that with negative spillovers no such results can be established. In fact, with SCD-C from below more productive types may find it optimal to choose strictly less items than less productive types.

FIGURE 2: THE POLICY FUNCTION OVER A ONE-DIMENSIONAL TYPESPACE



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 decision set. The single crossing in type assumption ensures that such groups exist. The resulting policy function $\mathcal{J}^\star(\cdot)$ changes its value only at each cutoff z_n, for $n = 1, 2, 3$.

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

3.4. The Generalized Squeezing Procedure

With heterogeneous agents, 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 two 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, i.e., the outputs of the bounding set functions evaluated for any $z \in Z$. The "auxiliary" set M collects items the algorithm has already considered but could not make progress on.

Whereas an application of the squeezing step from Section 3.2 only updated 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-tuples 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

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

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

The mapping S^b is defined as

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

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

We use an example, to show how to use the generalized squeezing step and connect to the original squeezing procedure. Consider a CDCP with a single dimension of heterogeneity and with a return function that satisfies SCD-T and SCD-C from above. A natural starting point for the algorithm, we set the initial 4-tuple as follows: $Z = \mathbf{Z}$, $\underline{\mathcal{L}}(\cdot) = \varnothing, \overline{\mathcal{L}}(\cdot) = L, \forall z \in Z$, and $M = \varnothing$. To apply the generalized squeezing step S^a to the corresponding 4-tuple, we first choose an undetermined item $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$. We

can identify the "cutoff" agent types $z^{\text{in}}, z^{\text{out}} \in Z$ which are exactly indifferent between including ℓ in $\overline{\mathcal{L}}$ and $\underline{\mathcal{L}}$ respectively:

$$0 = D_\ell \pi(\overline{\mathcal{L}}; z^{\text{in}}) \qquad\qquad 0 = D_\ell \pi(\underline{\mathcal{L}}; z^{\text{out}}) \,.$$

These cutoff types divide the original region Z into *up to* three subregions.[19]

For all types $z \in Z$ with $z < z^{\text{out}}$,

$$0 = D_\ell \pi(\underline{\mathcal{L}}; z^{\text{out}}) \quad \Rightarrow \quad 0 \geq D_\ell \pi(\mathcal{L}^\star(z); z^{\text{out}}) \quad \Rightarrow \quad 0 \geq D_\ell \pi(\mathcal{L}^\star(z); z) \,.$$

The first inequality follows from SCD-C, since $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star(z)$ for all $z \in Z$. The second inequality follows from SCD-T. We can then conclude that all types $z \in Z$ below z^{out} exclude j from their optimal decision set. Likewise, for all $z \in Z$ with $z > z^{\text{in}}$,

$$0 = D_\ell \pi(\overline{\mathcal{L}}; z^{\text{in}}) \quad \Rightarrow \quad 0 \leq D_\ell \pi(\mathcal{L}^\star(z); z^{\text{in}}) \quad \Rightarrow \quad 0 \leq D_\ell \pi(\mathcal{L}^\star(z); z)$$

Again, the first inequality follows from SCD-C, since $\mathcal{L}^\star(z) \subseteq \overline{\mathcal{L}}$ for all $z \in Z$, and the second from SCD-T. Given SCD-T and SCD-C, it is easy to verify that $z^{\text{out}} \leq z^{\text{in}}$.

The two cutoffs we identified create three new 4-tuples. After dividing Z into three new subregions according to the two cutoffs z^{out} and z^{in}, we update each subregion's bounding and auxiliary sets with the new information. For all $z > z^{in}$, ℓ is optimally included, so the subset $\underline{\mathcal{L}}$ includes ℓ. For all $z < z^{out}$, l is optimally excluded, so the superset $\overline{\mathcal{L}}$ excludes ℓ. For all $z \in (z^{out}, z^{in})$, we cannot conclude that ℓ is either optimally included or excluded. For this intermediate region, we instead add ℓ to M to encode the information that, given the current bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ on Z, ℓ remains undetermined. We have now created three new subregions from the original subregion Z, each with at least one of the original $\underline{\mathcal{L}}$, $\overline{\mathcal{L}}$, or M updated.[20]

The next theorem establishes that if a CDCP's underlying return function exhibits SCD-C

[19] It is irrelevant whether the type mass function $f(\cdot)$ assigns positive values to the cutoff values $z^{\text{in}}, z^{\text{out}} \in Z$.

[20] Returning to Theorem 2, we can verify the above example in single dimensional type space precisely corresponds to one application of the generalized squeezing step. As a technical detail, it is possible for either z^{out} or z^{in} to be outside of Z. In these cases, the generalized squeezing step will return one or two subintervals instead of three. The formal definitions of the generalized squeezing steps allow for this possibility.

and SCD-T, then each application of the generalized squeezing step updates the bounding sets without excluding items that are part of the optimal decision for any subregion Z of the type space.

Theorem 2. *Consider a CDCP as defined in equation (7) 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*.*

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* $\mathbf{z} \in Z$*.*

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* $\mathbf{z} \in Z$*.*

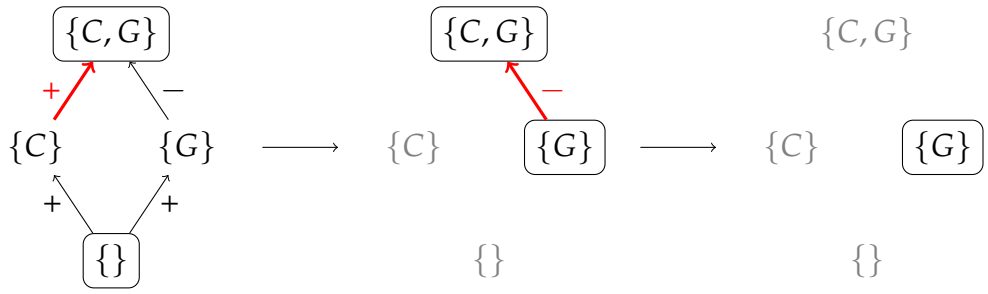
Conditional on the appropriate SCD-C restriction, each the recursive application of mappings S^a *and* S^b *converges in* $O(n)$ *time.*

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 generalized branching procedure in the Appendix.

A Simple Example To elucidate our solution concepts, we consider a simple example of our multinationals framework where there are 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 plant location 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

FIGURE 3: A SIMPLE EXAMPLE: LOCATING PLANTS IN GERMANY AND CANADA



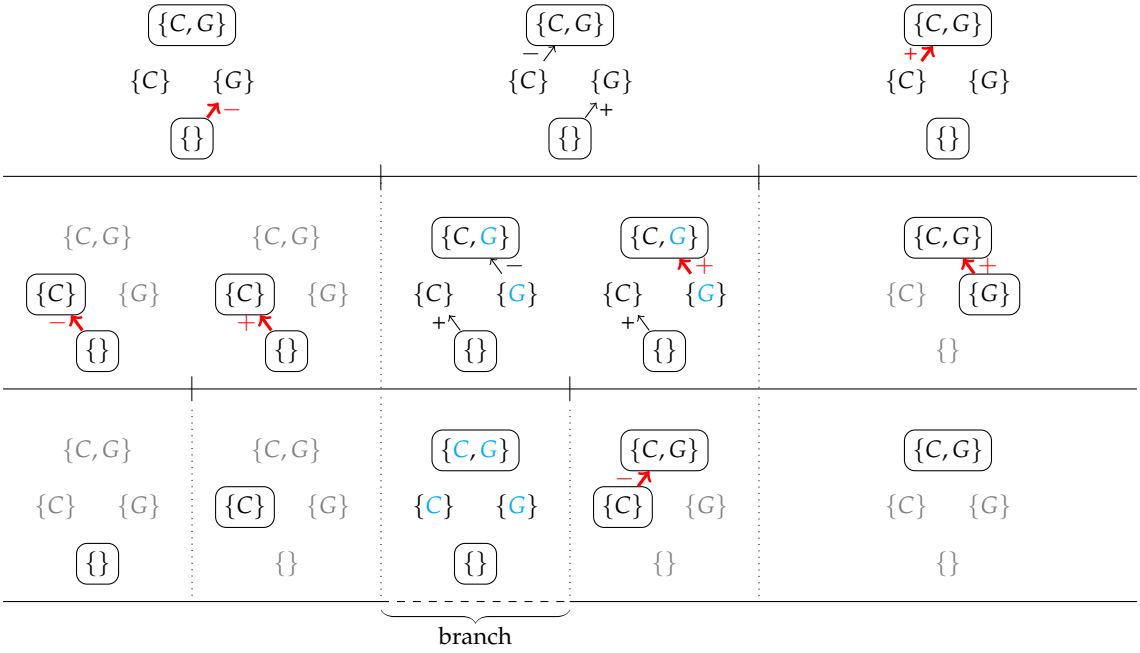
Notes: Applying the squeezing step three times to the simple Germany-Canada choice space. After each application, the choice space shrinks as countries are removed from the superset or added to the subset (denoted with a rectangle). Ultimately, one decision set remains.

$[\{\}, \{G, C\}]$; $\mathcal{L}^\star \in [\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, as required. Beginning with the subset, $\underline{\mathcal{L}} = \{\}$, we consider the return to adding Canada and Germany, separately. Since both countries have positive marginal values and SCD-C from above holds, we cannot discard either location as not optimal. 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 the Canadian plant is removed. The optimal decision set hence includes a German plant 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 decision set with certainty. Since the marginal value of a Canadian plant is negative in this context, we conclude that the firm optimally only opens a plant in Germany, so that $\mathcal{L}^\star = \{G\}$.

Figure 4 illustrates the use of the generalized squeezing procedure to solve for the policy function in the same context.Each row of Figure 4 illustrates one application of the generalized squeezing step. In the first row, the entire interval $\mathbf{Z}$ shares an identical bounding pair $[\{\}, \{C, G\}]$ indicated with rectangles. As before, the signs along the arrows indicate the marginal value of adding a given location to the bounding set. Red arrows imply that an interval's bounding pair can be updated.

FIGURE 4: AN EXAMPLE OF THE GENERALIZED SQUEEZING PROCEDURE



Notes: A possible recursive sequence from the generalized squeezing procedure.

In the first row, we consider adding a German plant. We identify two cutoff productivities, marked with vertical ticks. For firm types below the first cutoff, adding a plant in Germany when there is no plant in Canada yields negative marginal benefit; these firms never open a plant in Germany. For firm types above the second cutoff, adding a plant in Germany while there is no plant in Canada yields positive marginal benefit; these firms always open a plant in Germany. For firms in the middle interval, no update is possible so G enters the auxiliary set M since no decision could be reached yet; we encode this update by representing Germany in blue. The middle row of Figure 4 shows the type space with updated bounding pairs.

Given the updated boundary sets, in the leftmost interval, we identify the cutoff productivity for which opening a plant in Canada has positive marginal value; for this interval there is no subregion of the type space for which not decision can be reached. In the rightmost interval, all types receive positive marginal value from opening a German plant. In the middle interval, we identify a cutoff above which firms receive positive marginal benefit from a Canadian plant when a German plant exists.

The third row of Figure 4 reflects updates from the second row. We have found the optimal decision set for intervals one, two, and five. Interval four has to be updated one more time, to reflect that these types optimally open a Canadian plant only. For interval three, both C and G are in the auxiliary set and there remain no countries to consider; the branching step has to be applied (not shown).

4. Establishing SCD-C and SCD-T and Closing the Model

In this section, we establish the SCD-C and SCD-T conditions in our framework of multinational production introduced in Section 2. We then introduce additional parametric assumptions in preparation for the quantitative application and close the model by specifying the full set of equilibrium conditions. Throughout the rest of the paper, we assume that the firm-specific productivity shifter does not differ across a firm's plants, i.e., that $z_\ell = z \; \forall \ell$. This allows us to provide sharper results and intuition and is the relevant case in most quantitative applications; the assumption is not essential for the results that follow.

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

Consider the framework presented in Section 2. In Appendix A, we show that the profit maximization problem of firm headquartered in location i with productivity z satisfies SCD-C from below as long as:

$$\underbrace{\frac{\mathrm{d}\ln q_{in}(\mathcal{L};z)}{\mathrm{d}\ln p_{in}(\mathcal{L};z)}\frac{\mathrm{d}\ln p_{in}(\mathcal{L};z)}{\mathrm{d}\ln c_{in}(\mathcal{L};z)}}_{\text{Demand Channel}} \geq \underbrace{1+\theta}_{\text{Supply Channel}}, \tag{9}$$

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

The expression in equation 9 provides sufficient statistics for the strength of positive complementarities working through the demand side (LHS) and the strength of the negative complementarities working through the supply side of the model (RHS). SCD-C from below, which corresponds to positive complementarities, holds precisely as long as

the strength of the demand channel exceeds that of the supply channel.

The left hand side shows that the strength of the demand side channel is given by the product of the demand and passthrough elasticities. Their product determines how much incremental marginal cost savings raise the firm's sales and hence how much the marginal cost savings associated with an additional plant raise the value of all other plants of the firm. Likewise, the relevant measure of the strength of the supply side complementarity is simply $1+\theta$, which is an inverse measure of how substitutable plants are in the firm's production function. If θ is high, the firm largely chooses supplying plants for each variety based on trade costs and productivities common across all varieties within a plant, which leaves plants more substitutable – so cannibalization among plants is stronger.

It turns out that the product of demand and pass-through elasticity also plays a central role in establishing SCD-T in our framework. In particular, in Appendix A, we establish that our framework exhibits SCD-T as long as

$$\frac{\mathrm{d}\ln q_{in}(\mathcal{L};z)}{\mathrm{d}\ln p_{in}(\mathcal{L};z)}\frac{\mathrm{d}\ln p_{in}(\mathcal{L};z)}{\mathrm{d}\ln c_{in}(\mathcal{L};z)} \geq 1. \tag{10}$$

Intuitively, the SCD-T condition is akin to requiring the cross-complementarity between plants and productivity to be positive. Equation (10) once again separates the condition for this to hold into a demand-side effect and a supply-side effect.The left hand side captures how the value of an individual plant changes with firm productivity through demand side channels. Since firm productivity, like additional plants, just appear as a marginal cost shifter, the strength of this channel is again summarized by same statistic as the between-plant complementarity on the demand side. On the cost side, additional plants and higher productivity act as substitutes, since both lower marginal cost. Since firm productivity and market potential are multiplicative in the case of single dimensional firm heterogeneity z (cf. equation 3), this effect is 1.

In Appendix OA.1, we present generalizations of the conditions in equations 9 and 10 for verifying SCD-C and SCD-T in version of our framework without the Fréchet assumption, i.e., for more general marginal cost functions $c_{in}(\mathcal{L};z)$. Our more general conditions nest, but are not limited to, both the Fréchet case presented here and the case of the multivariate Pareto presented in Arkolakis et al. (2018).

4.2. The Equilibrium System in the Quantitative Framework

For our quantitative exercise, we follow functional form choices from a recent literature in international trade and multinational production (Helpman et al. (2004); Tintelnot (2017); Arkolakis et al. (2018)) and specialize the general demand function in equation 1 to that implied by a CES demand system with elasticity σ. In the CES case, the aggregate demand and price shifters take the familiar forms of $P_n^{1-\sigma} = \sum_i \int_{\Omega_i} (p_{in}(\omega))^{1-\sigma} d\omega$ and $Q_n^{\frac{\sigma-1}{\sigma}} = \sum_i \int_{\Omega_i} (q_{in}(\omega))^{\frac{\sigma-1}{\sigma}} d\omega$. The resulting pricing rule is a special case of the pricing rule in equation 4:

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

which features a constant markup over marginal costs.[21] Our condition for SCD-C from below in equation 4 collapses to $\sigma \geq 1 + \theta$, which reflects the fact that markups are constant in the CES case so that pass-through is 1 and the demand elasticity is simply σ. The condition for SCD-T collapses to $\sigma \geq 1$. Therefore, in this special case, SCD-T is implied by the SCD-C condition.

We finally turn to aggregation over firms and the determination of aggregate variables in general equilibrium. To show the equilibrium conditions, we make use of the policy function notation for location choices. Since productivities and all bilateral costs differ arbitrarily across (i, ℓ, n) pairs, firm policy functions are origin-country specific. We denote by ${\mathcal{L}^\star}_i(z)$ the location choice of a firm with productivity z headquartered in location i.

In each country of origin i, firms have to pay a fixed labor cost f_i^e to establish a headquarter. After entry, they draw their productivity from the distribution $G_i(z)$. Firms continue to enter until expected profits are zero. As a result, in equilibrium, the total mass of entrants in each origin country i, M_i, satisfies the free entry condition

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

where $\mathbb{1}_{i\ell}(z)$ is an indicator variable for whether a firm with productivity z from origin i optimally opens a production location in country ℓ and where X_n denotes total final goods spending in destination n.

[21] We present the case with the Pollak demand system in Appendix OA.3.

However, not all entrants end up producing positive quantities of output, since low productivity firms do not find it profitable to pay the fixed costs of opening a production plant. In particular, there is a cutoff productivity for each location that determines the minimum firm productivity needed to profitably establish at least one production plant. The cutoff, $\underline{z}_i$ is pinned down by the following zero profit condition:

$$\pi_i(\mathcal{L}_i^\star(\tilde{z}_i);\tilde{z}_i) = 0. \tag{12}$$

Given the mass of entrants in location i, M_i, pinned down by equation 11, the mass of final product varieties offered by firms headquartered in location i in equilibrium is then simply $\Omega_i = (1 - G_i(\tilde{z}_i))M_i$. Note that, in a slight abuse of notation, we use Ω_i both to denote the mass of final products by firms from location i and the set of their products.

Price indices in each destination market n aggregate the individual prices of all goods offered in that country. We rewrite this condition taking into account that firms from the same origin country i with the same productivity behave identically:

$$P_n^{1-\sigma} = \sum_i M_i \int_z p_{in}\left(\mathcal{L}^\star{}_i(z);z\right)^{1-\sigma} \mathrm{d}G_i(z). \tag{13}$$

This expression incorporates the fact that firms will choose the plant network $\mathcal{L}^\star{}_i(z)$, which in turn determines their pricing decision together with their idiosyncratic productivity z.

Next, the labor market must clear in each production location ℓ. There are two sources of labor demand: variable labor requirements from any firms with a production location in ℓ, as well as the fixed costs of entry and plant establishment. The following condition summarizes these and equates it to labor supply:[22]

$$\begin{aligned} w_\ell L_\ell = {} & \frac{\sigma-1}{\sigma}\sum_{i,n} X_n M_i \int_z \mathbb{1}_{i\ell}(z)\frac{\gamma_{i\ell} w_\ell \tau_{\ell n}}{\sum_{\ell'\in\mathcal{L}^\star{}_i(z)}\gamma_{i\ell'} w_{\ell'}\tau_{\ell' n}}\left(\frac{p_{in}\left(\mathcal{L}^\star{}_i(z);z\right)}{P_n}\right)^{1-\sigma}\mathrm{d}G_i(z) \\ & + M_\ell w_\ell f_\ell^e + w_\ell \sum_i M_i f_{i\ell}\int_z \mathbb{1}_{i\ell}(z)\mathrm{d}G_i(z). \end{aligned} \tag{14}$$

[22] 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 $\frac{\sigma-1}{\sigma}$ of total sales under CES demand. Of these total production costs, production location ℓ accounts for the share $\frac{\gamma_{i\ell} w_\ell \tau_{\ell n}}{\sum_{\ell'\in\mathcal{L}^\star{}_i(z)}\gamma_{i\ell'} w_{\ell'}\tau_{\ell' n}}$ out of the firm's network $\mathcal{L}$.

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

$$X_n = w_n L_n \tag{15}$$

Having laid out the key aggregate conditions governing the spatial economy, we now define general equilibrium.

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 given in (6)
2. Given the policy functions, the aggregates satisfy equation (11), (12), (13), (14), and (15).

For the quantitative implementation, we assume that the firm productivity distribution 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.

5.1. Data

Our calibration relies on data on trade flows and multinational production (MP) sales from a data set created by Alviarez (2019) with information on 9 economic sectors in 32 OECD and non-OECD countries.[23]

[23]The countries are Australia, Austria, Belgium, Bulgaria, Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Italy, Japan, Latvia, Lithuania, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russian Federation, Slovakia, Spain, Sweden, Turkey, Ukraine, United Kingdom, and the United States.

For each country and sector the data set contains information on the number of establishments and total sales of establishments by origin country, (e.g., German manufacturing plants and their sales in France) and trade flows by sector, origin, and destination country. Each observation in the dataset is an origin-destination-sector triplet averaged over the period from 2003 to 2012.[24] The data set also contains information on bilateral tariffs and distance as well as contiguity and common language indicators. We supplement these data with information from the Penn world tables, which contain real GDP per capita and total employment for all countries in our sample.

To illustrate the patterns in the data, Table 1 shows a Poisson Maximum Likelihood (PPML) regression of trade shares, inward MP sales shares, and inward MP plant shares on origin-sector and destination-sector fixed effects and bilateral distance and tariffs in the full Alviarez (2019) data.[25] Interestingly MP and trade flows correlate differently with tariffs. All else equal, higher tariffs are associated with lower trade flows, as has been shown in a large number of data sets and as consistent with economic theory. However, MP flows and tariffs are positively correlated, highlighting that MP may act as a substitute for trade between countries: instead of shipping goods from A to B, firms from A can set up a plant in B to supply both B and nearby countries in a cheaper way. Since shipping between A and B is more expensive with high tariffs, such MP activity becomes more attractive.[26] In Appendix, Section OA.5 we present robustness checks to the regressions in Table 1; the regression coefficients are robust across the different specifications.

[24] To construct the dataset, Alviarez (2019) combines data from the OECD, the Eurostat Foreign Affiliate Statistics database, and the the Bureau of Economic Analysis.

[25] In particular, we run three Poisson Pseudo Maximum Likelihood (Silva and Tenreyro (2006)) regression in the data to inform θ and potentially other paramters in our model:

$$y_{ijs}^{x} = \exp\left(\alpha^{x} + \beta_{d}^{x} \log d_{ij} + \beta_{t}^{x} \log\left(1 + t_{ijs}\right) + \delta_{x}' X_{ij} + \varkappa_{is} + \varkappa_{js}\right) + \epsilon_{ijs}^{x}$$

where d_{ij} indexed distance between countries i and j, t_{ijs} are sector-specific tariffs, and X_{ij} is a vector of gravity controls. $\varkappa_{is}$ and $\varkappa_{js}$ are origin-sector and destination-sector specific fixed effects. We estimate the equation for three outcome variables indexed by x: trade shares ($x = T$), MP shares ($x = MP$), and plants ($x = P$).

[26] One of the most robust findings is that increases in tariffs lead to a substitution of trade with Foreign Direct Investment, as in Brainard (1993); Antràs and Yeaple (2014). This finding is consistent with our findings of a positive coefficient on the regression of MP and a negative on trade flows.

TABLE 1: Regressions in the Data

	Trade Shares	MP Shares	Plants
Log Distance	-0.908***	-0.642***	-0.394***
	(0.0288)	(0.0666)	(0.0719)
Log Tariffs	-4.274*	8.176***	2.004**
	(1.970)	(1.499)	(0.616)
Observations	8369	8160	8013

Notes: The table presents the estimated coefficients on tariffs and distance from running a PPML gravity regression in the Alviarez (2019) data. The regression include origin-sector and destination-sector fixed effects. and dummies for common language, for geographical contiguity, and for colonial ties. The differences in the number of observations across specifications reflect different numbers of non-missing observations for the respective outcome variable. Marginal effects; Standard errors in parentheses., $\sym{*} \(p<0.05\), \sym{**} \(p<0.01\), \sym{***} \(p<0.001\)$.

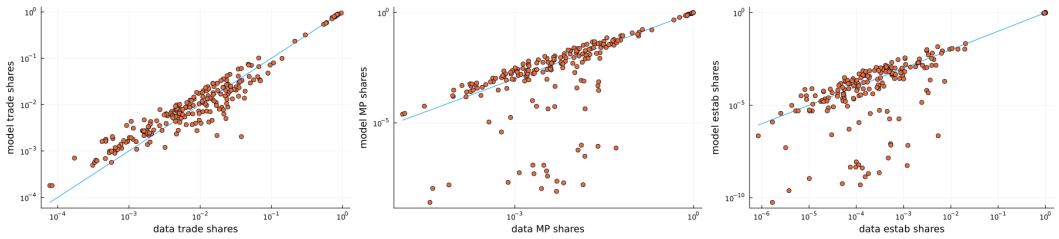
5.2. Calibration Strategy

Our calibration data set includes the 15 largest countries in terms of GDP from the Alviarez (2019) dataset. To calibrate our model, we target data on GDP per capita and total outward multinational sales for each country, as well as the full matrix of trade, MP, and plant shares from the data. The calibration of the Fréchet shape parameter θ is still in progress.

Trade Costs, MP Costs, Fixed Costs and Country Productivity In the theory, the bilateral trade, MP, and fixed location costs are matrices with $N \times N$ entries, where N is the number of countries. To calibrate these objects, we first normalize their diagonal values to 1 so that $\tau_{ii} = \gamma_{ii} = f_{ii} = 1 \ \forall i \in N$ without loss of generality. We then calibrate the off-diagonal values to match the full matrices of trade shares, inward MP shares, and plant shares among our 15 countries observed in the data. In contrast to many models that yield closed-form expressions for gravity equations (e.g., Costinot and Rodríguez-Clare (2014), Redding and Rossi-Hansberg (2017), Allen et al. (2020)) the lack of closed form gravity in our model implies that we cannot necessarily match these shares exactly.

Figure 5 graphs the model-implied trade and MP shares against those in the data in the

FIGURE 5: Trade Shares, Inward MP Sales Shares, and Inward MP Plant Shares in Data and Model



Notes: The left-panel plots the model predicted trade shares against the trade shares in the data. The middle panel plots the model predicted inward MP sales shares against the same object in the data. The right panel plots the model predicted inward MP plant shares against the same object in the data.

calibrated model. The fit is good for larger trade, inward MP sales , and and inward MPplant shares (correlation in levels of about 0.98 and above) but not as good for smaller shares, where the correlation drops substantially.

In addition to the cost matrices, we calibrate two country-specific productivity terms: the scale parameter of the firm Pareto disitribution, $\underline{z}_i$, which acts as location-of-headquarter productivity shifter, and the scale of the plant-variety Fréchet distribution, T_ℓ, which acts as a location-of-production productivity shifter. We choose $\{\underline{z}_i\}_i$ to exactly match the total amount of foreign affiliate sales conducted by firms headquartered in i relative to the US, and $\{T_\ell\}_\ell$ for each country to exactly match the observed GDP per capita for each country from the Penn World Tables.

Trade elasticity and Fréchet Elasticity For each variety and each plant, firms draw a productivity from a Fréchet distribution with shape parameter θ. The inverse of the shape parameter is a measure of the strength of comparative advantage differences among a firm's plants. If $1/\theta$ is low, plants do not differ much in this random productivity component and firms largely choose among plants on the basis of trade costs, tariffs, and country productivity differences. If $1/\theta$ is large, the productivity of each plant differs widely across varieties so that the firm tends to choose supplying plants for a given variety mainly based on the variety's idiosyncratic productivity component rather than based on tariffs associated with the plant's location. As a result, the higher $1/\theta$, the less

MP flows should respond to tariffs and distance costs. We are still developing a strategy to calibrate θ. For now we set $\theta = 2.82$.

Other Parameters The Pareto distribution of firm productivity is governed by the dispersion parameter ξ. The model allows for a closed form expression for the Pareto tale which allows us to choose ξ to match the shape of the Pareto tail estimated in other studies. In particular, given σ, we choose ξ so that $1.65 = \xi/(\sigma - 1)$ holds in the model, in line with the estimates presented in Arkolakis (2010).

The elasticity of substitution between final good varieties σ governs the ease with which consumers move consumption spending between firms' outputs. If σ is high consumer expenditure responds to differences in relative prices across firms. We set $\sigma = 5.5$ which is within the range of estimates of Broda and Weinstein (2006).

Lastly, the parameter η, which governs the elasticity of substitution between a firms own inputs, turns out to be irrelevant for solving general equilibrium and conducting counterfactuals, similarly to in Antras et al. (2017). As a result, we remain agnostic of its value.

5.3. Computational Performance of the Algorithm

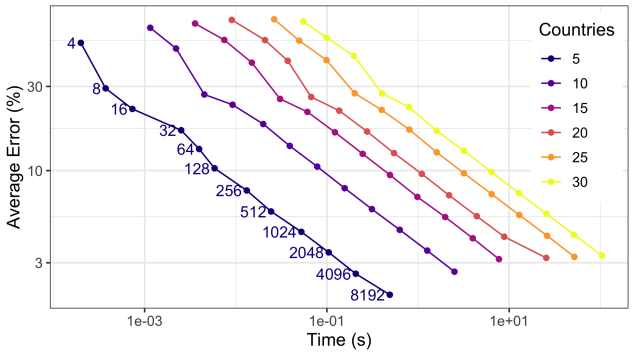
In this section, we provide a number of speed tests to showcase the computational performance of our algorithm to solve CDCPs and aggregate them across heterogeneous agents.

First, we compare how long it takes to solve for the full policy function mapping firm productivity to the optimal set of plants. To demonstrate differences between solution methods we compare three ways of solving for the policy function: (1) Discretize firm productivity with 1024 grid points, and solve the CDCP at each grid point by comparing the profit associated with any possible combination of plants and choosing the maximum ("Naive"), (2) Discretize firm productivity with 1024 grid points, and use our squeezing and branching algorithm above to solve the CDCP at each grid point ("Single"), (3) use our generalized squeezing and branching methods to solve for the full policy function without discretization ("Policy"). Figure 6 presents the algorithm's run time in seconds for different numbers of countries for each of three methods. For 15 countries the policy

FIGURE 6: Comparing Methods of Computing the Firm Policy Function

Countries	(1) Naive	(2) Single	(3) Policy
5	0.11	0.05	0.00
10	9.15	0.30	0.14
15	540.06	0.96	1.53
20	≈days	3.12	11.59
25	≈mnths	6.27	38.59
30	≈yrs	12.22	129.16

Performance by Method



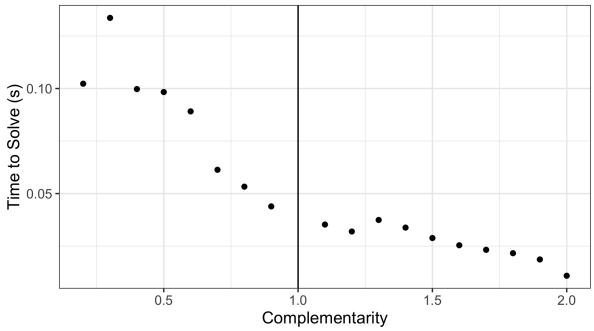
Precision-Time Tradeoff

Notes: The table on the left illustrates the computational time for computing the firm policy function with a grid of 1024 points with the simple brute force, the method of Theorem 1 (and the refinement) and the method of Theorem 2. The figure on the right illustrates the approximation error in computing the policy function with the "single" approach, for different amounts of countries. To measure approximation error, we first compute the percentage deviation for each trilateral flow $X_{i\ell n}$ when using the discretized policy function compared to the true policy function. The approximation error is the average percentage deviation across all triplets (i, ℓ, n).

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 increases at a polynomial –not exponential– rate. Notice that the policy function is a bit slower, but to provide a proper comparison of the computational times we need to take into account that Naive and Single are approximations, while the policy is computed to (machine error) precision.

We study the trade-off between approximation error and computing time using the "single" approach. The right panel in Figure 6 graphs computational time against this error, with different number of grid points and total number of countries. For example, with four grid points and five countries, computation time is low but discretization error is above 30%. The precision of the "Single" is no better than 4-5% with 1024 grid points and improves to about 2% with 8192 points but at the cost of about 10 times higher computation 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.

FIGURE 7: Complementarities and Computational Speed



Notes: The figure illustrates the computational time that it takes for our algorithm to solve the firm policy function with 15 countries depending on the degree of complementarity, defined as the ratio of $(\sigma - 1)/\theta$. High negative complementarities (ratio below 1), take more time.

Finally, Figure 7 shows how computational time depends on the strength and direction of complementarities. We solve the model several times for 15 countries using the "Policy" method above but vary the the ratio of parameters that determines the strength and direction of the complementarities between plants in our model, $(\sigma - 1)/\theta$. We find that it is generally faster to solve for the policy function in the presence of positive complementarities (ratio above 1). For negative complementarities computational time increases at a manageable rate, roughly tripling the computational time when the ratio is 0.5 versus when the ratio is 2.

6. Counterfactual Exercises

In this final section we use the model to simulate two scenarios of two recent major political events with economic consequences, specifically relating to multinational firm operations. The first is the exit of Great Britain from the EU, commonly referred to as Brexit, and the second the US-EU sanctions on Russia following the Russian invasion of Ukraine, which we refer to as Rusexit.

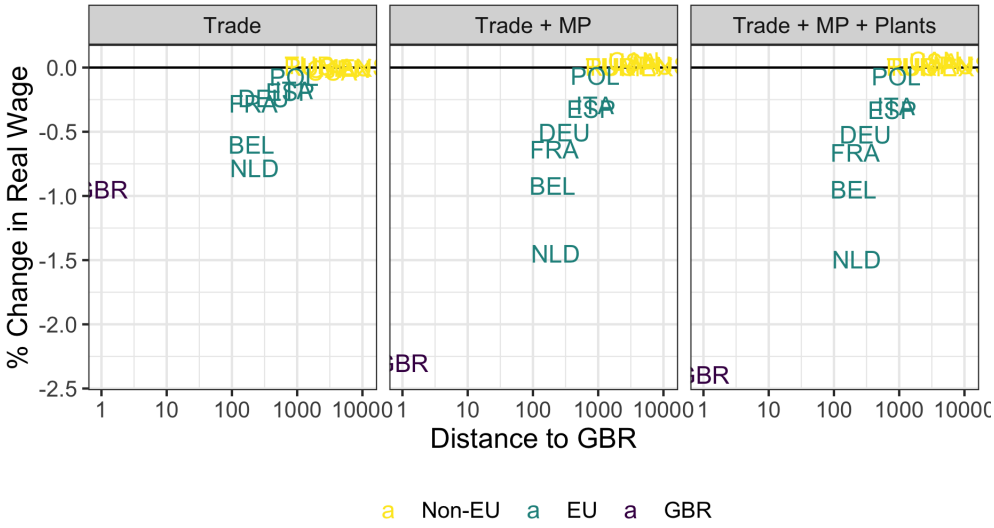
6.1. Exit of Great Britain from the European Union

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 plant cost of production location. In the counterfactual exercise below, we progressively increase these frictions between Great Britain and EU member countries by 10%.

Figure 8 summarizes the impact of these changes on welfare as captured by real wages in the three scenarios, plotted against distances 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 real wage drops by 1%. 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 and, thus, bilateral distance. Great Britain's closest trading allies, such as Belgium, France, and the Netherlands, suffer the most.

Once variable MP costs also rise, the impact on welfare grows starker. As we explain below, in reaction to the increased variable MP costs between Great Britain and EU members, many firms based in EU member states no longer operate plants in Great Britain.The resulting lower labor demand in Great Britain exerts downward pressure on wages. At the same time, since many firms no longer produce in Great Britain,s their goods can reach the country only via trade, putting upwards pressure on the price index in Great Britain. A similar effect occurs in EU member states as firms based in Great

FIGURE 8: Brexit's Impact on Real and Nominal Wages



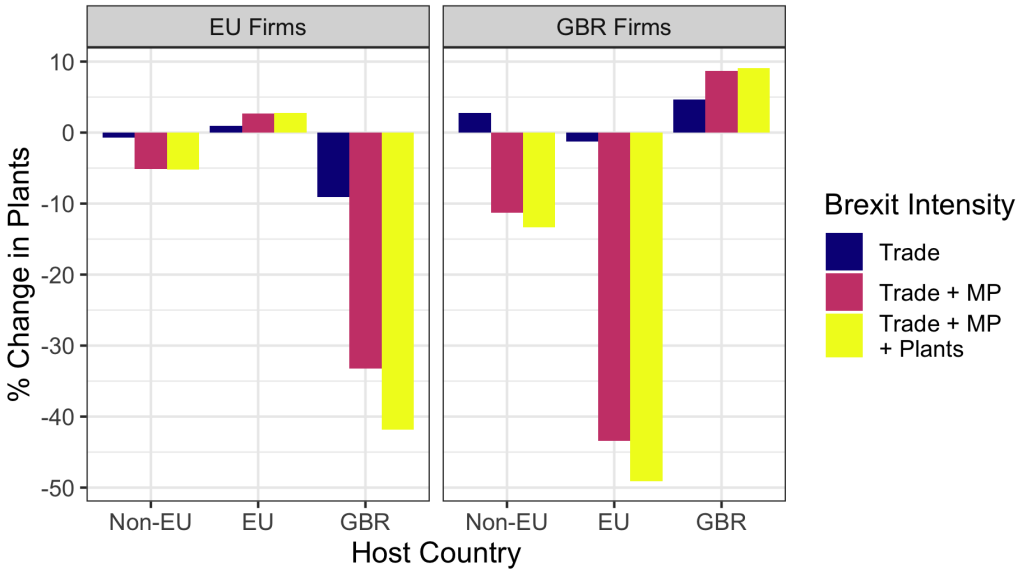
Notes: The effect of increasing frictions between Great Britain and EU member states (in green). In the first panel, only trade costs increase. The second panel adds increases to variable MP costs. The last adds increases to fixed costs of plant establishment.

Britain withdraw production from EU member states. Overall, nominal wages increase in these countries, but real wages remain constant or rise only a little. Effects of geography are evident as the Brexit consequences are stronger to countries in geographic proximity to Great Britain, reflecting the fact that MP costs are related to distance. Adding on increases to the fixed costs of production between Great Britain and EU member states, in the third panel, accentuates this pattern.

The effects through the exit of foreign establishments set our framework apart from predictions of traditional trade models, in which wage changes are accommodated with price index changes of similar magnitudes leading to muted effects on real wages, as pointed out by Dekle et al. (2007). Instead, in our exercise, the combined effect is a drop in real wages that is more than proportional to the drop in nominal wages.

To gain more insight into these shifts in production patterns, we plot the mass of operating plants in Figure 9. The first panel displays the change in the mass of plants opened by firms headquartered in an EU member state, in host countries that are non-members, members, or in Great Britain. The increase in trade costs leads to a modest drop in

FIGURE 9: Openings and Closings of Plants by Sender and Host Country

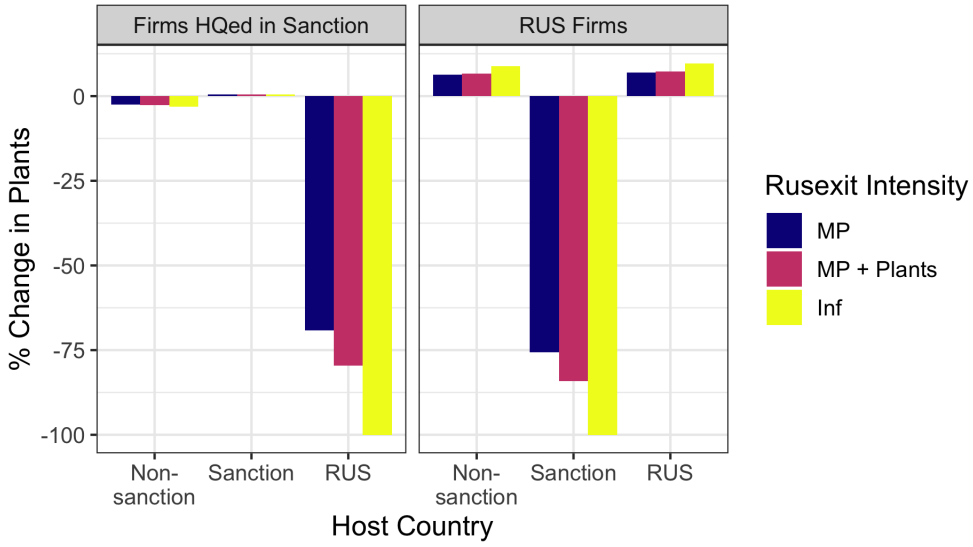


Notes: The first panel of the figure shows the mass plants operated by firms headquartered in the EU present in non-EU countries, EU countries, and in Great Britain in the baseline and in the two counterfactual scenarios. The second panel shows the mass plants operated by firms headquartered in Great Britain present in non-EU countries, EU countries, and in Great Britain in the two counterfactual scenarios compared to the baseline.

the mass of EU-operated affiliates in Great Britain, accompanied by a slight increase in the mass of EU-operated affiliates in EU member states. This effect reflects the export platform motive of plant location in our framework. As it becomes difficult to serve nearby EU member states from the plant in Great Britain, EU firms elect to relocate their production to countries that remain EU member states. The rise of MP costs, both iceberg and fixed, substantially magnifies the latter effect and leads EU firms to close even more plants in Great Britain while opening plants in other countries.

The right panel shows the mass of plants operated by firms based in Great Britain. In the baseline, British firms open many plants in EU member states. With the new frictions, they open substantially fewer plants in EU member states. In contrast, the number of plants British firms operate in Great Britain increases substantially. However, the withdrawal of British firms from EU markets implies that they have higher costs, thereby significantly impacting British consumers' real wages as we discusse above.

FIGURE 10



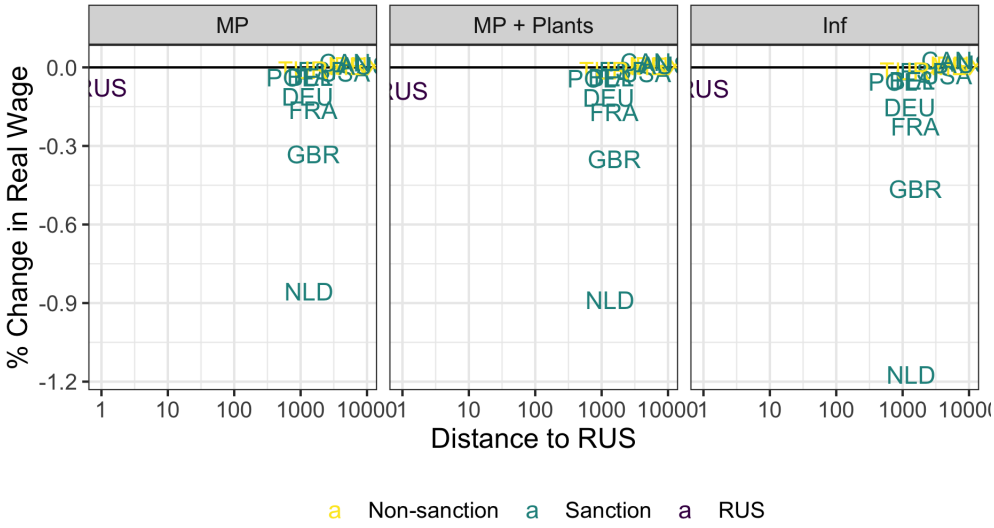
Notes: *Notes*: The first panel of the figure shows the change in the mass of plants operated in Russia by firms headquartered in countries that placed sanctions to Russia, for only iceberg MP costs increase, adding increases to fixed costs of plant establishment, and setting MP costs to infinity. The second panel shows the same changes but for firms headquartered in Russia.

6.2. 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 USA, Canada, European Union countries, and Australia. In particular, we progressively impose a 30% increase in the bilateral MP iceberg cost, an additional 30% increase in the MP fixed cost, and finally infinite iceberg and MP costswhich altogether prohibit MP between the sanctioning countries and Russia.

To begin, Figure 10 summarizes the effect on bilateral plant establishment of the policy. The panel on the left describes the changes in the mass of plants operated by firms headquartered in countries that placed sanctions. As the sanctions are progressively introduced, there is a large drop in multinational presence from these firms in Russia. Overall, bilateral plant establishments with Russia fell in most countries that imposed sanctions. With only increases in iceberg MP costs, the mass of plants in Russia operated

FIGURE 11



Notes: The effect of increasing frictions between the Russian Federation and countries imposing sanctions (in green). In the first panel, only iceberg MP costs increase. The second panel adds increases to fixed costs of plant establishment. The last panel sets MP costs to infinity, completely prohibiting multinational production between the Russian Federation and countries imposing sanctions.

by firms based in countries placing sanctions drops by 70%. [27]Increasing the fixed costs of plant establishment also modestly amplifies this effect. As multinational productions costs continue to rise to infinity, all previous plants operated by these firms in Russia must close, so that the mass of plants in Russia drops by 100%. By comparison, there is little change in their production presence among other countries.

The panel on the right describes the production presence of firms based in Russia. The main effect of the sanctions is to decrease the mass of plants operated by Russian firms in countries placing sanctions. Increase, the mass of plants increases in both countries not implementing sanctions as well as in Russia itself. The increase of plants in countries not placing sanctions is in line with the export platform motive of FDI. On the other hand, the rise in plants operated domestically reflects the loosening of competition in both the Russian labor market as well as the Russian goods market, as firms previously operating plants there withdraw.

[27]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 11 draws out the effects of these changes in plant locations for real wages. Russia does not suffer large losses, since both inward and outward MP are a relatively small fraction of Russian GDP. However, countries that impose sanctions, like the Netherlandsand Great Britain, have non-negligible losses of around 0.9% and 0.3% respectively. MP sales with Russia uniformly decline by about 70% for the countries imposing sanctions in our simulations. The Netherlands and Great Britain are particularly negatively affected given their geographic proximity to many other countries that also impose sanctions and the ensuing geographic network effects as they the European Union collectively withdraws production from Russia. USA, Canada, and Australia also impose sanctions but are geographically remote from the European countries. Countries that do not impose sanctions are affected little as they are geographically remote to Europe. Adding the fixed increase does not make a big difference as the outcomes of imposing either frictions are very correlated and increasing the costs to infinity only slightly exacerbates those patterns.

7. Conclusion

Multinational location decisions are combinatorial as they involve negative or positive complementarities among the many potential choices. We introduce a new methodology to solve combinatorial discrete choices with negative or positive complementarities and put it to use to solve and estimate a quantitative model of multinational location and production decisions. In a series of counterfactuals that resemble the exit of Great Britain from the EU we find that increases in multinational production and fixed costs lead to large plant exit of British plants from EU countries and EU plants from Britain and large losses from Brexit for Great Britain and many European countries. Similarly, simulating the USA-EU-Australia and Russian Federation firm sanction war leads to large plant exit but only the welfare of a handful of countries is meaningfully impacted. Geographic distance and geographic network effects determine the impact of those shocks in such counterfactual, signifying the importance of geography in our analysis.

References

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.

AMITI, M., O. ITSKHOKI, AND J. KONINGS (2019): "International shocks, variable markups, and domestic prices," *The Review of Economic Studies*, 86, 2356–2402.

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," in *Handbook of international economics*, Elsevier, vol. 4, 55–130.

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. RODRÍGUEZ-CLARE (2019): "The elusive pro-competitive effects of trade," *The Review of Economic Studies*, 86, 46–80.

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.

BRAINARD, S. L. (1993): "An empirical assessment of the proximity-concentration tradeoff between multinational sales and trade," .

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

BRYAN, G. AND M. MORTEN (2019): "The aggregate productivity effects of internal migration: Evidence from indonesia," *Journal of Political Economy*, 127, 2229–2268.

CORNUÉJOLS, G., G. L. NEMHAUSER, AND L. A. WOLSEY (1983): "The uncapacitated facility location problem," Tech. rep., DTIC Document.

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

COSTINOT, A. AND A. RODRÍGUEZ-CLARE (2014): "Trade theory with numbers: Quantifying the consequences of globalization," in *Handbook of international economics*, Elsevier, vol. 4, 197–261.

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

DEKLE, R., J. EATON, AND S. KORTUM (2007): "Unbalanced trade," *American Economic Review*, 97, 351–355.

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.

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.

GUADAGNI, P. M. AND J. D. LITTLE (1983): "A logit model of brand choice calibrated on scanner data," *Marketing science*, 2, 203–238.

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.

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

JAKOB, K. AND P. M. PRUZAN (1983): "The simple plant location problem: survey and synthesis," *European journal of operational research*, 12, 36–81.

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," .

KLENOW, P. J. AND J. L. WILLIS (2016): "Real rigidities and nominal price changes," *Economica*, 83, 443–472.

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.

LIND, N. AND N. RAMONDO (2018): "Trade with correlation," Tech. rep., National Bureau of Economic Research.

MATSUYAMA, K. AND P. USHCHEV (2017): "Beyond CES: three alternative classes of flexible homothetic demand systems," *Global Poverty Research Lab Working Paper*.

MCFADDEN, D. (1973): "Conditional logit analysis of qualitative choice behavior," .

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 (2020): "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.

REDDING, S. J. AND E. ROSSI-HANSBERG (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

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.

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, econometrics, and an application to automobile demand*, vol. 10, MIT press.

VERTER, V. (2011): "Uncapacitated and capacitated facility location problems," in *Foundations of location analysis*, Springer, 25–37.

YANG, C. (2020): "Location choices of multi-plant oligopolists: Theory and evidence from the cement industry," .

YEAPLE, S. R. (2013): "The multinational firm," *Annu. Rev. Econ.*, 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

A.1. Proofs of Main Theorems

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

The proof of Theorem 1 proceeds as follows.

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 $L^\star$ and that the lower bounding set contains no values that are not in $L^\star$. 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 $\mathcal{L}^\star \subseteq \overline{\mathcal{L}}^{(1)}$. So we established 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 $\ell \in \mathcal{L}^\star$ optimally. 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 $\ell \in \overline{\mathcal{L}}^{(k+1)}$.

So we established 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.

Finally, the squeezing procedure must complete in under $|L|$ iterations. Each iteration, if no new items are fixed, then the procedure has converged. As a result, it must be that each iteration fixes at least one item if the procedure continues. Thus, the maximal number of iterations is achieved when exactly one item is fixed each time, with all $|L|$ items eventually being fixed. □

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 $\Lambda_\ell^c(\underline{\mathcal{L}}_0)$ 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. □

A.2. Nesting Structure of the Policy Function

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

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

Proof. The proof proceeds by contradiction. Suppose $\mathcal{L}^\star(z_1) \not\subseteq \mathcal{L}^\star(z_2)$. Define $\mathcal{L}^o \equiv \mathcal{L}^\star(z_1) \setminus \mathcal{L}^\star(z_2)$, which has cardinality $N \in (0, \infty)$ by assumption. Further, let $\mathcal{L}^i \equiv \mathcal{L}^\star(z_1) \cap \mathcal{L}^\star(z_2)$ so that $\mathcal{L}^\star(z_1) = \mathcal{L}^i \cup \mathcal{L}^o$.

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

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

where the first line derives from z_1 optimality, the second from single-dimensional SCD-T, and the third SCD-C from above, and the fourth from recognizing that $\ell \notin \mathcal{L}^\star(z_2)$. [28]This

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

argument proves the claim for $N = 1$.

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

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

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

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

where the last inequality makes use of the inductive assumption. Thus we show the claim holds for $n + 1$ if it holds for n.

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

The next proposition extends Proposition 1 to the case of multidimensional types $\mathbf{z}$, under a slighlty stricter condition on type-plant complementarity in place of SCD-T.

Proposition 2 (Nested policy function on multidimensional type space)**.** *Consider a CDCP (c.f. Definition 1) for which the return function satisfies SCD-C from below and for each component i of the type vector $\mathbf{z}$,*

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

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

Proof. aThe proof proceeds by contradiction. Suppose $\mathcal{L}^\star(\mathbf{z}_1) \not\subseteq \mathcal{L}^\star(\mathbf{z}_2)$ so that $\mathcal{L}^o \equiv \mathcal{L}^\star(\mathbf{z}_1) \setminus \mathcal{L}^\star(\mathbf{z}_2) \neq \emptyset$. Let $\mathcal{L}^i \equiv \mathcal{L}^\star(\mathbf{z}_1) \cap \mathcal{L}^\star(\mathbf{z}_2)$ so that $\mathcal{L}^\star(\mathbf{z}_1) = \mathcal{L}^i \cup \mathcal{L}^o$. Observe that, for any convex combination of the two types $\mathbf{z}_1 + \theta(\mathbf{z}_2 - \mathbf{z}_1)$, the marginal value of item

set.

j in decision set $\mathcal{L}$ for this type can be expressed using the line integral

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

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

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

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

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

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

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

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

In particular, observe that the condition replacing SCD-T is stricter than the sufficiency

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

A.3. Sufficient Conditions for SCD-C and SCD-T

In this section, we discuss the single crossing difference conditions from the main text. We begin by clarifying the relationship between sub- and super-modularity with SCD-C. We show that the first pair of conditions are sufficient for the second; we then show that the second are sufficient for the first when the choice space is finite. A short discussion follows outlining a counterexample for the case of an infinite choice space. Finally, we verify the sufficient condition for SCD-T provided in the main text.

Proposition 3 (Sufficiency of sub- and super-modularity)**.** *In this first proposition, we show that submodularity and supermodularity are sufficient to ensure SCD-C. Fix the agent type* $\mathbf{z}$ *and consider the return function* π*.*

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

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

Proof. Since $\mathbf{z}$ is fixed during this proof, we suppress in the following for notational brevity.

Begin with a submodular mapping π. Then, by the definition of submodularity, for any sets A, B, it is the case that

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

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

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

$$\pi(\mathcal{L}_1) - \pi(\mathcal{L}_1 \setminus \{\ell\}) \geq \pi(\mathcal{L}_2) - \pi(\mathcal{L}_2 \setminus \{\ell\}) \qquad \text{if } \ell \in \mathcal{L}_1, \text{ so } \ell \in \mathcal{L}_2$$

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

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

Then, it is easy to see that applying the submodularity condition implies SCD-C from above.

Now, suppose π is supermodular. Then, for any sets A, B, it is the case that

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

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

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

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

In this second proposition, we show that, conditional on a finite choice space L, the SCD-C property implies sub- or super-modularity.

Proposition 4 (Sufficiency of SCD-C with finite choice space)**.** *Fix an agent type* **z***and consider the return function π. Let A and B be arbitrary sets so that $A \setminus (A \cap B)$ is finite.*

If π exhibits SCD-C from above, then

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

If π exhibits SCD-C from below, then

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

Proof. Since the agent type $\mathbf{z}$ is fixed during this proof, we suppress it notational brevity in what follows.

Let $\tilde{A}$ and $\tilde{B}$ be arbitrary sets where $\tilde{A} \setminus (\tilde{A} \cap \tilde{B})$ is finite. First consider the SCD-C from above property. Define

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

Then showing that

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

is equivalent to showing that:

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

The proof proceeds inductively on the cardinality of A. When A is empty, then equation A.1 holds with equality.

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

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

while from SCD-C from above,

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

Combining the two expressions together yields

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

which confirms equation A.1 for sets A of cardinality $n+1$. The inductive proof estab-

lishes that equation A.1 holds for all A of finite size.

Next, consider SCD-C from below. The argument follows a similar structure. Now, it is equivalent to show that

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

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

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

while from SCD-C from below,

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

Combining the two expressions together yields

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

which confirms equation A.2 for sets A of cardinality $n+1$. The inductive proof establishes that equation A.2 holds for all A of finite size. □

When $A \setminus (A \cap B)$ is not finite, then SCD-C is not sufficient to guarantee the supermodularity or submodularity properties. As a simple counterexample, suppose the return π of a decision set S is defined

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

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

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

The intuition is simple: since we integrate over the a decision set S for its return, any

singular element ℓ is measure zero and has no effect on the decision set's overall return. The return function therefore satisfies SCD-C (of both forms). Now consider $A = [0, 2]$ and $B = [1, 3]$. It is easy to see that

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

so, in this case,

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

Then, when $\alpha > 1$, the return function obeys SCD-C from above but violates submodularity. Likewise, when $\alpha \in (0, 1)$, the return function obeys SCD-C from below but violates supermodularity.

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

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

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

Proposition 5 (Sufficient condition for SCD-T)**.** *Fix an item ℓ and $\mathcal{L}$. Let the entries of $\mathbf{z}$ be indexed by i, so that z_i is the ith coordinate of $\mathbf{z}$. Suppose*

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

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

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

construct the point $\tilde{\mathbf{z}}$ where

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

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

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

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

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

We similarly construct a piece-wise linear path from $\mathbf{z}'$ to $\tilde{\mathbf{z}}$ that lies in $\Lambda_\ell(\mathcal{L})$. Joining these paths together at $\tilde{\mathbf{z}}$, we have constructed a path from $\mathbf{z}$ to $\mathbf{z}'$ that remains in $\Lambda_\ell(\mathcal{L})$. Since $\mathbf{z}$ and $\mathbf{z}'$ were any arbitrary members of $\Lambda_\ell(\mathcal{L})$, we have shown that it is path-connected. Showing $\Lambda_\ell^c(\mathcal{L})$ is path-connected follows a similar argument. Now suppose $\mathbf{z}$ and $\mathbf{z}'$ are contained in $\Lambda_\ell^c(\mathcal{L})$. We construct $\tilde{\mathbf{z}}$ in this case as

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

We then construct the paths from $\mathbf{z}$ to $\tilde{\mathbf{z}}$ and from $\mathbf{z}'$ to $\tilde{\mathbf{z}}$ in the same way. Both lie in $\Lambda_\ell^c(\mathcal{L})$ by the same logic. □

ONLINE APPENDIX FOR

COMBINATORIAL DISCRETE CHOICE: A QUANTITATIVE MODEL OF MULTINATIONAL LOCATION DECISIONS

BY COSTAS ARKOLAKIS, FABIAN ECKERT, AND ROWAN SHI

FOR ONLINE PUBLICATION ONLY

OA.1. SCD-C and SCD-T in the General Framework

In this section, we provide sufficient conditions under which the SCD-C and SCD-T assumptions hold in a general class of models nesting our theoretical framework in Section 2. Throughout this section, we assume that the firm-specific productivity takes a scalar form so that $z_\ell(\omega) = z(\omega) \equiv z \ \forall \ell \in L$.

OA.1.1. Assumptions on the Cost Function

Consider a firm of productivity $z \in \mathbb{R}^+$ headquartered in country i with a plant network $\mathcal{L}$ and the unit cost of delivering its final good to a destination market n,$c_{in}(\mathcal{L}, z)$. In Section 2, we presented a particular microstructure for $c_{in}(\mathcal{L}, z)$ (cf. Equation 3). Instead, in this section, we directly impose structure on $c_{in}(\mathcal{L}, z)$ that implies that the SCD-C and SCD-T conditions hold.

Assumption 1. *The cost function of a firm headquartered in country iwith productivity z in destination n can be written as* $c_{in}(\mathcal{L}, z) = g(\Theta_{in}(\mathcal{L}), z)$ *where* $g : \mathbb{R}^2 \to \mathbb{R}^+$ *and* $\Theta : \mathscr{P} \to \mathbb{R}$ *and*

1.The "index function" Θ *is monotonically increasing in the sense that* $\mathcal{L} \subseteq \mathcal{L}' \to \Theta_{in}(\mathcal{L}) \leq \Theta_{in}(\mathcal{L}')$ *for all* $\mathcal{L}, \mathcal{L}' \subseteq L$ *(if it is decreasing, redefine* $\tilde{\Theta_{in}}(\mathcal{L}) = -\Theta_{in}(\mathcal{L})$*); and* Θ_{in} *features no interdependencies between elements of* $\mathcal{L}$*, i.e.,* $D_\ell \Theta_{in} \equiv \Theta_{in}(\mathcal{L} \cup \{\ell\}) - \Theta_{in}(\mathcal{L})$ *does not depend on* $\mathcal{L}$ *for all* $\ell \in L$ *and any* $\mathcal{L} \subseteq L$*.*

2.The "cost function" g is monotonically decreasing in firm productivity (if it's increasing redefine $\tilde{z} \equiv -z$*) and in the value of the index function, i.e.,* $\partial g / \partial z \leq 0$ *and* $\partial g / \partial \Theta_{in} \leq 0$ *hold for all* $\mathcal{L} \in L$ *and* $z \in \mathbb{R}^+$*.*

Note that the cost function $c_{in}(\mathcal{L}, z)$ that we obtained from the microstructure imposed in Section 2 satisfies Assumption 1 with the production potential Θ_{in} corresponding to the index function that maps a plant strategy into a scalar: the contribution of an individual plant ℓ to the index is independent of which other plant are included $\mathcal{L}$. In Section 2, the cost function is the following non-linear function of the index function and firm productivity: $g(\Theta_{in}(\mathcal{L}), z) = \tilde{\Gamma} \frac{1}{z} \left[\Theta_{in}(\mathcal{L})\right]^{-1/\theta}$.

OA.1.2. Establishing SCD-C and SCD-T

Throughout this Section, we assume that the cost function satisfies Assumption 1 and firms confront the general demand system introduced in Section 2.[29] We show that for cost functions that satisfy Assumption 1 there exist easy-to-verify condition to check whether the firm's problem satisfies SCD-C and SCD-T.

SCD-C We first derive a condition for SCD-C.

First notice that given Assumption 1, the marginal contribution of a given plant to the value of the index Θ_{in} is independent of other plants. As a result, without loss of generality, we can write Θ_{in} as the simple sum of the marginal effects of each location on the index:

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

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

$$D_k \pi_i(\mathcal{L}; z) = \sum_n \left[\pi_{in}(c_{in}(\mathcal{L} \cup k; z)) - \pi_{in}(c_{in}(\mathcal{L} \cap k; z))\right] - f_{ik} \tag{OA.1}$$

where $\pi_{in}(c(\mathcal{L} \cup k; z)) = \pi_{in}(g(\Theta_{in}, z))$denotes the variable profits of the firm headquartered in i with productivity z in destination market n conditional on production potential $\Theta_{in}(\mathcal{L})$ and f_{ik} is the fixed cost of setting up the location in k. The marginal value represents the gain in variable profits from increasing production potential by ξ_{ikn} which is offset, in part, by the additional fixed costs incurred.The SCD-C condition requires that the expression in equation OA.1 only crosses zero once. To show this, it is sufficient to show the marginal value is monotonic, i.e. , given any $\mathcal{L}_1 \subseteq \mathcal{L}_2 \subseteq L$, the marginal value of any given item k is bigger (smaller) at $\mathcal{L}_2$ than for $\mathcal{L}_1$ for SCD-C from below (above). Comparing this marginal value across two decision sets and making use

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

of the fundamental theorem of calculus we have,

$$D_k\pi_i(\mathcal{L}_2;z) - D_k\pi_i(\mathcal{L}_1;z) = \sum_n \int_0^{\xi_{ikn}} \left[\frac{\partial \pi_{in}}{\partial \Theta_{in}} \left(y + \sum_{\ell \in \mathcal{L}_2} \xi_{i\ell n}, z \right) - \frac{\partial \pi_{in}}{\partial \Theta_{in}} \left(y + \sum_{\ell \in \mathcal{L}_1} \xi_{i\ell n}, z \right) \right] \mathrm{d}y$$

$$= \sum_n \int_{\Theta(\mathcal{L}_1)}^{\Theta(\mathcal{L}_2)} \frac{\partial}{\partial x} \left[\int_0^{\xi_{ikn}} \frac{\partial \pi_{in}}{\partial \Theta_{in}}(x+y,z)\mathrm{d}y \right] \mathrm{d}x = \sum_n \int_{\Theta(\mathcal{L}_1)}^{\Theta(\mathcal{L}_2)} \left[\int_0^{\xi_{ikn}} \frac{\partial^2 \pi_{in}}{\partial \Theta_{in}^2}(x+y,z)\mathrm{d}y \right] \mathrm{d}x$$

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

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

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

The second derivative of the destination-specific profit function with respect to the production is given by:

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

where the elasticity of the derivative of the profit function is

$$\varepsilon_{\pi'} = -\frac{\frac{\partial \pi_{in}^2(c)}{\partial c^2}}{\frac{\partial \pi_{in}(c)}{\partial c}} c^2 = \underbrace{\varepsilon_q(p)}_{\text{demand elasticity}} \underbrace{\frac{\mathrm{d} \ln p}{\mathrm{d} \ln c}}_{\text{passthrough}},$$

where the last equality uses the generic expression for firm profits, $\pi(c) = q(p(c))p(c) - q(p(c))c$. The sign of $\frac{\partial^2 \pi_{in}}{\partial \Theta_{in}^2}$ is determined by the component in square brackets since the term pre-multiplying them is positive ($\pi_1' \leq 0$). All together, the firm's profit function satisfies SCD-C from below if the term in square brackets is positive, i.e., if $\varepsilon_{\pi'}$ exceeds

$\frac{\varepsilon_{g'_1}}{\varepsilon_g}$, and SCD-C from above if the inequality is reversed..

On the supply side, the right-hand-side of the equation, plants can exert positive or negative complementarities on each other. With negative complementarities, an additional plant reduces marginal costs by less the more plants the firm already operates ("cannibalization") with positive complementarities the opposite is the case.

On the demand side, the left-hand-side of the equation, there are two easily interpretable statistics: the elasticity of demand and the rate of passthrough of cost shocks to prices. Plants always exert *positive* complementarities on each other through both channels. First, when the firm produces large quantities, additional marginal cost savings are more valuable because they are applied to the firm's entire production volume. The force of this effect is given by the elasticity of demand, which links the firm price to firm scale. The second source of complementarity derives from the fact that firms may only partially pass cost savings onto consumers, absorbing the rest in higher markups. This force is naturally governed by the firm's passthrough rate.

SCD-T A similar argument holds for SCD-T in our context, when the productivity is Hicks-neutral:

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

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

$$\pi_{in}\left(\frac{\nu\left(\mathcal{L}\cup\ell\right)}{z}\right)-\pi_{in}\left(\frac{\nu(\mathcal{L})}{z}\right)-f_{i\ell}=\int_{\frac{\nu(\mathcal{L})}{z}}^{\frac{\nu(\mathcal{L}\cup\ell)}{z}}\pi'_{in}(c)\mathrm{d}c-f_{i\ell}=\int_{\nu(\mathcal{L})}^{\nu(\mathcal{L}\cup\ell)}\pi'\left(\frac{\nu}{z}\right)\frac{1}{z}\mathrm{d}\nu-f_{i\ell}$$

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

$$\pi_{in}\left(\frac{\nu\left(\mathcal{L}\cup\ell\right)}{z}\right)-\pi_{in}\left(\frac{\nu(\mathcal{L})}{z}\right)-f_{i\ell}=\int_{\nu(\mathcal{L}\cup\ell)}^{\nu(\mathcal{L})}\left[-\pi'_{in}\left(\frac{\nu}{z}\right)\right]\frac{1}{z}\mathrm{d}\nu-f_{i\ell}$$

where $\pi'_{in}\leq 0$ since variable profits decrease in marginal cost. SCD-T requires that this marginal value crosses zero at (at most) one productivity value. It is sufficient to show that the marginal value is monotonic in productivity. Comparing the marginal value at

two values of productivity $z_1 \leq z_2$,

$$\int_{\nu(\mathcal{L}\cup\ell)}^{\nu(\mathcal{L})} \left[-\pi'_{in}\left(\frac{\nu}{z_2}\right)\right]\frac{1}{z_2}\mathrm{d}\nu - \int_{\nu(\mathcal{L}\cup\ell)}^{\nu(\mathcal{L})} \left[-\pi'_{in}\left(\frac{\nu}{z_1}\right)\right]\frac{1}{z_1}\mathrm{d}\nu$$

$$= \int_{z_1}^{z_2} \frac{\partial}{\partial z}\left\{\int_{\nu(\mathcal{L}\cup\ell)}^{\nu(\mathcal{L})} \left[-\pi'_{in}\left(\frac{\nu}{z}\right)\right]\frac{1}{z}\mathrm{d}\nu\right\}\mathrm{d}z = \int_{z_1}^{z_2}\int_{\nu(\mathcal{L}\cup\ell)}^{\nu(\mathcal{L})} \left[-\pi'_{in}\left(\frac{\nu}{z}\right) - \frac{\nu}{z}\pi''_{in}\left(\frac{\nu}{z}\right)\right]\left(-\frac{1}{z^2}\right)\mathrm{d}\nu\mathrm{d}z$$

$$= \int_{z_1}^{z_2}\int_{\nu(\mathcal{L}\cup\ell)}^{\nu(\mathcal{L})} \left[-\frac{1}{z^2}\pi'_{in}\left(\frac{\nu}{z}\right)\right]\left[\underbrace{-\frac{\pi''_{in}\left(\frac{v}{z}\right)}{\pi'_{in}\left(\frac{v}{z}\right)}\frac{\nu}{z}}_{\equiv\varepsilon_{\pi'}} - 1\right]\mathrm{d}\nu\mathrm{d}z \int_{z_1}^{z_2}\int_{\nu(\mathcal{L}\cup\ell)}^{\nu(\mathcal{L})} \left[-\frac{1}{z^2}\pi'_{in}\left(\frac{\nu}{z}\right)\right][\varepsilon_{\pi'} - 1]\,\mathrm{d}\nu\mathrm{d}z$$

so it is sufficient to show this difference is positive. Note that the first square bracketed term is positive. Then, it is sufficient for $\varepsilon_{\pi'} \geq 1$ for SCD-T to hold. Using again $\pi(c) = q(p(c))p(c) - q(p(c))g$, we get that:

$$\varepsilon_{\pi'} = \varepsilon_q(p)\frac{\mathrm{d}\ln p}{\mathrm{d}\ln c}$$

So that SCD-T holds as long as

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

It is straightforward to extend these proofs to the case of a non-Hicks Neutral productivity shifter.

OA.2. The Branching Procedure

In this section, we provide the details about the basic branching procedure to maximize the objective function of a single agent and the generalized branching procedure to solve for the policy function mapping firm types to optimal strategies in the context of a model in which several heterogeneous firms are solving CDCPs.

OA.2.1. The Basic Branching Procedure

At the heart of the branching procedure is a "branching step" applied to a CDCP for which the squeezing procedure has converged. The branching step takes an undetermined item ℓ such that $\ell \in \overline{\mathcal{L}}^{(K)} \setminus \underline{\mathcal{L}}^{(K)}$ and forms two subproblems, or "branches:" one in which ℓ is included in $\mathcal{L}^\star$ (i.e., added to $\underline{\mathcal{L}}$) and one in which it is excluded from $\mathcal{L}^\star$ (i.e., excluded from $\overline{\mathcal{L}}$).[30] The two fixed points resulting from applying the squeezing procedure to the bounding sets of each subproblem are the optimal decision sets *conditional* on the assumed inclusion or exclusion of ℓ. The optimal decision set of the original CDCP is then the conditional optimal decision set that yields the higher value of π.

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

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

Definition 8 (Branching step)**.** Given bounding sets $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$, select some element $\ell \in \overline{\mathcal{L}} \setminus \underline{\mathcal{L}}$.

The mapping B^a is given by

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

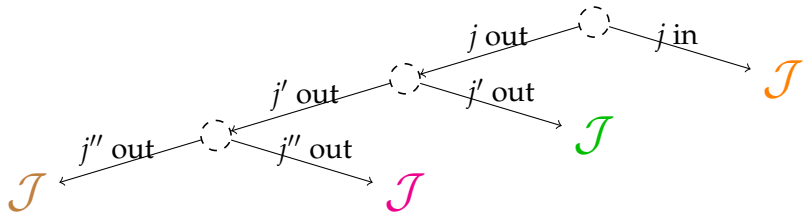
The mapping B^b is given by

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

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

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

FIGURE OA.1: EXAMPLE PATH OF THE BRANCHING PROCEDURE



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

each branch.[31]

Suppose the return function exhibits SCD-C from above.[32] Given an initial bounding pair $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ with $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}$, the globally converged result $B^{a(K)}([\underline{\mathcal{L}}, \overline{\mathcal{L}}])$ is a collection of branch-specific optimal decision sets. The cardinality of the set is the number of branches. Among these conditionally optimal decision sets, the one yielding the highest value of π is the optimal decision set solving the original CDCP. Note that contrary to the squeezing procedure, the branching procedure *always* identifies the optimal decision set.

For exposition, suppose the return function satisfies SCD-C, and consider Figure OA.1 which shows an example of a tree created by the branching procedure. It starts with a bounding set pair for which the squeezing procedure has converged, but there still remain undetermined items. One of these, ℓ, is selected. Two branches based on this item are formed. The left hand branch corresponds to the subproblem where ℓ is presumed to be excluded from the optimal decision set, while the right hand branch corresponds to the subproblem where ℓ is presumed to be included. The squeezing procedure is reapplied in each branch. On the right hand branch, the squeezing procedure delivers a bounding pair where $\underline{\mathcal{L}} = \overline{\mathcal{L}}$, yielding the orange $\mathcal{L}$. This decision set is optimal conditional on the requirement that ℓ must be included. On the other hand, convergence of the squeezing procedure in the left hand branch does not deliver an optimal decision set. The returned

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

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

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

To summarize the branching procedure, consider a CDCP as defined in equation 7 and let the bounding pair $[\underline{\mathcal{L}}, \overline{\mathcal{L}}]$ be such that $\underline{\mathcal{L}} \subseteq \mathcal{L}^\star \subseteq \overline{\mathcal{L}}$.

Then, if π exhibits SCD-C from above,

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

while if π exhibits SCD-C from below,

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

OA.2.2. The Generalized Branching Procedure

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

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

The mapping B^a is given by

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

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

The mapping B^b is given by

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

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

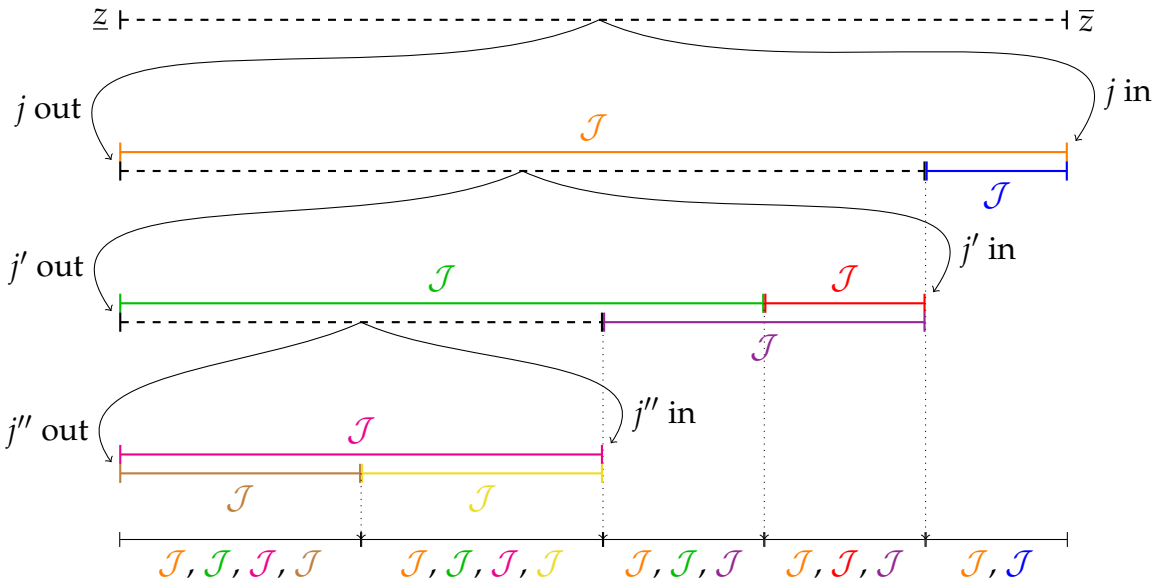
Given an initial 4-tuple $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]$ and an undetermined item $\ell \in M$, the generalized branching step creates two branches or subproblems. The first supposes that ℓ is included in the optimal decision set, while the second supposes that it is excluded. To each of these two subproblems we apply the generalized squeezing procedure until global convergence obtaining a collection of 4-tuples which exhaustively partition the original region Z.[33] Each branch may now contain different partitions of the original type space.

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

For illustration, Figure OA.2 depicts the process of applying the generalized branching procedure to an initial 4-tuple $[(\underline{\mathcal{L}}, \overline{\mathcal{L}}, M), Z]$. The initial 4-tuple specifies lower and upper bound sets $(\underline{\mathcal{L}}, \overline{\mathcal{L}})$ over the entire dotted interval Z between $\underline{z}$ and $\overline{z}$. Applying the generalized branching step once, the problem is divided into two subproblems: one corresponding to requiring item ℓ to be excluded, and the other requiring that item ℓ be included. In the subproblem on the right branch, the squeezing procedure identifies the single (orange) decision set that is optimal for the whole interval conditional on including ℓ. On the left branch, ℓ is excluded. In this case, convergence from the squeezing procedure delivers a policy function only for the highest types $z \in Z$, identifying the (blue) optimal decision set. Undetermined elements remain for the lower types of the range. The branching step is thus reapplied for this subsection of the original interval, selecting a second undetermined element, ℓ'. This procedure repeats until no undetermined elements remain in any of the branches for any type $z \in Z$.

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

FIGURE OA.2: THE GENERALIZED BRANCHING PROCEDURE: AN EXAMPLE OUTCOME



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

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

OA.3. The Pollak Demand System

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

The class of demand systems introduced by Pollak (1971) has become popular in the literature studying variable markups.[34] Consider the following demand function from the Pollak class which is frequently used in quantitative applications (e.g., Arkolakis et al. (2019)) and first appeared in Klenow and Willis (2016):

$$q_n(\omega) = Q_n D\left(p_n(\omega)/P_n\right) = \left(p_n(\omega)/P_n^\star\right)^{-\sigma} + \gamma, \quad \text{where} \quad Q_n = 1 \tag{OA.3}$$

where $\gamma < 0$. The demand in equation OA.3 has a variety of appealing features. First, it features a choke price, $P_n^\star$, which implies that entry into each destination market n is guaranteed only for the firms with low enough marginal costs, $c_n \leq P_n^\star$. In other words, a firm does not necessarily serve all countries, but self-selects into export markets consistent with the data (see also Arkolakis et al. (2019)). Second, asymptotically, the elasticity of demand is constant, which allows the model to fit the Pareto-size tails of firm distribution for the largest firms and exporters, a key feature of the data (see Arkolakis (2016); Amiti et al. (2019)). Finally, under very general conditions, it implies that markups

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

increase with firm size, a salient finding of recent investigations on the relationship of firm size and firm markups (see De Loecker et al. (2016)).

The pricing rule implied by the demand function in equation OA.3 is given by

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

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

Finally, the statistic necessary to establish whether SCD-C holds for a given CDCP (cf. Section OA.1.2) takes the convenient form:

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

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

OA.4. The Simple Plant Location Problem

The Simple Plant Location Problem ("SPLP") or Uncapacitated Plant Location Problem ("UPLP") refers to a general class of operation research problems concerned with the optimal choice of plants on a topography so as to minimize transportation costs while serving a set of spatially distributed demand points. The SPLP is a canonical problem in Operations Research (e.g., p-Center and p-Median problems) (see Krarup and Pruzan (1983), Cornuéjols et al. (1983), Owen and Daskin (1998), and Verter (2011) for surveys).

While the problem has been shown to be NP-hard, there are many heuristic solution methods. The algorithm we provide can be regarded as a heuristic solution method tailored to the structure of the type of plant location problems that occur in economics.

In this section, we show that the SPLP can be expressed as a CDCP as in Definition1 that satisfies the SCD-C property and can hence be solved using our methods.

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

$$\begin{aligned} \min_{\{\lambda_{\ell n}\}_{\ell \in L, n \in N}} & \sum_{\ell \in L} \sum_{n \in N} c_{\ell n} \lambda_{\ell n} + \sum_{\ell \in L} f_\ell \theta_\ell \\ \text{subject to } & \sum_{\ell \in L} \lambda_{\ell n} = 1 \ \forall n \\ & \theta_\ell \geq \lambda_{\ell n} \ \forall \ell, n \\ & \lambda_{\ell n}, \theta_\ell \in \{0,1\} \ \forall \ell, n \end{aligned} \tag{OA.4}$$

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

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

$$\pi_\ell(\mathcal{L}) = \mathbb{1}\left(\ell \in \mathcal{L}\right) \left[\sum_{n \in N} c'_{\ell n} \mathbb{1}\left(\max_{k \in \mathcal{L}} c'_{kn} \leq c'_{\ell n} \right) - f_\ell \right] \tag{OA.5}$$

The firms overall profit is the sum of the profits of its producing plants, i.e., $\pi(\mathcal{L}) =$

$\sum_{\ell \in \mathcal{L}} \pi_\ell(\mathcal{L})$. The problem in equation OA.4 can then be written as

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

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

Next, we show that the firm's objective satisfies the SCD-C property. In fact, it satisfies the stronger "submodularity" property which is sufficient for SCD-C from above. From the perspective of the firm, adding an additional plant will always weakly decrease the number of demand points served by any previously existing plant. As a result, the individual plant profit function in equation (OA.5) is weakly decreasing in the total number of operating plants. To see this formally, consider the following Lemma:

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

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

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

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

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

But then since inequalities are closed under addition:

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

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

OA.5. Additional Empirical Results

Table OA.1 presents robustness results for the regression run in Section 5. In each panel, there are three outcome variables: bilateral trade shares, bilateral MP shares, and bilateral MP plant stocks. We run all regressions across 32 countries and 9 tradable sectors and include origin-sector and destination-sector fixed effects. Panel A replicates the regression from Table 1 in Section 5. Panels B-D of Table OA.1 provide robustness checks on the results in Panel A by re-running it with different sets of controls. Relative to Panel A, Panel B removes the controls for language, geographical contiguity, and colonial ties. Panel C includes the same controls as Panel A but also a common currency dummy and s dummy for the existence of a regional trade agreement. Relative to Panel A, Panel D adds an additional regressor: a control for shared religion. With the exception of the tariff coefficient plant regression in Panel D, all specifications agree in terms of the sign and magnitude of the estimated coefficients.

Table OA.2 shows the relationship between the calibrated fixed costs, MP costs, and trade costs terms and distance. In the first two regressions, we drop pairs of countries for which observe no MP activity in the data since we infer infinite MP costs for them. We drop also all own country costs.

Figure OA.3 shows a histogram of our calibrated fixed, MP and trade costs. We drop the own-costs which are normalized to 1 for all three types of costs. We also drop country pairs for which calibrated MP costs are infinity, since MP flows in the data are zero.

TABLE OA.1: Robustness Regressions

	Trade Shares	MP Shares	Plants
	PANEL A		
Log Distance	-0.908***	-0.642***	-0.394***
	(0.0288)	(0.0666)	(0.0719)
Log Tariffs	-4.274*	8.176***	2.004**
	(1.970)	(1.499)	(0.616)
	PANEL B		
Log Distance	-1.036***	-0.735***	-0.591***
	(0.0460)	(0.0576)	(0.0877)
Log Tariffs	-5.672*	10.12***	4.155***
	(2.386)	(1.294)	(0.798)
	PANEL C		
Log Distance	-0.861***	-0.647***	-0.417***
	(0.0366)	(0.0700)	(0.0686)
Log Tariffs	-4.102*	8.094***	1.342*
	(1.898)	(1.550)	(0.574)
	PANEL D		
Log Distance	-0.849***	-0.615***	-0.399***
	(0.0424)	(0.0674)	(0.0656)
Log Tariffs	-5.151*	6.976***	-0.565
	(2.241)	(1.525)	(0.691)
Observations	8369	8160	8013

Marginal effects; Standard errors in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table presents the estimated coefficients on tariffs and distance from running a PPML gravity regression in the Alviarez (2019) data. All regressions include origin-sector and destination-sector fixed effects. Panel A includes dummies for common language, for geographical contiguity, and for colonial ties. Panel B includes no additional controls. Panel C includes the same controls as Panel A but also a common currency dummy and a dummy for the existence of a regional trade agreement. Relative to Panel A, Panel D adds an additional regressor: a control for shared religion. Standard errors are clustered at the sector level.

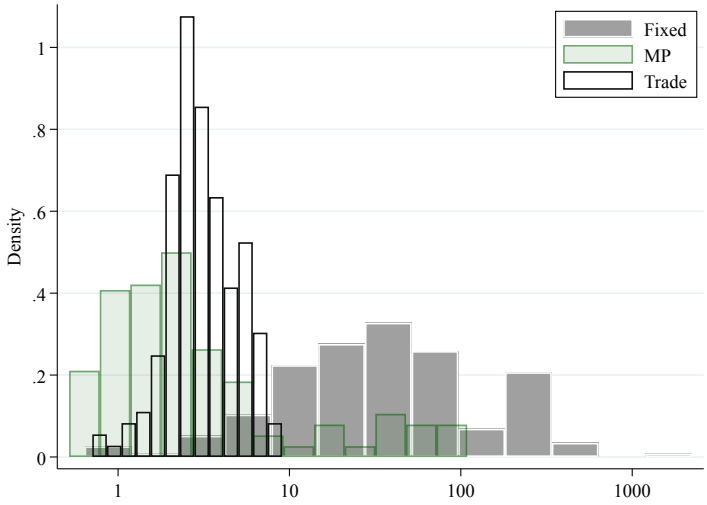
TABLE OA.2: Calibrated Costs and Distance

	Fixed Costs	MP Costs	Trade Costs
Log Distance	-0.0378	0.0451	0.183***
	(0.0796)	(0.0274)	(0.0183)
Common Language	-0.0167	-0.178**	-0.0750
	(0.156)	(0.0539)	(0.0388)
Contiguity	0.212	-0.0888	-0.0493
	(0.186)	(0.0642)	(0.0471)
	185	185	210

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The table shows the log of the calibrated fixed costs, MP costs, and trade costs terms regressed on log distance., a common language indicator, and a contiguous border indicator For fixed costs and MP costs we drop country pairs with zero MP for which we infer infinite MP costs and fixed costs of MP. We also drop own country pairs.

FIGURE OA.3: Histogram of Fixed, MP, and Trade Costs



Notes: The figure shows a histogram of our calibrated fixed, MP and trade costs. We drop the own-costs which are normalized to 1 for all three types of costs. We also drop country pairs for which calibrated MP costs are infinity, since MP flows in the data are zero.