Combinatorial Discrete Choice: Multinational Companies, Plant Fixed Costs, and Markups*

Costas Arkolakis
Yale University

Fabian Eckert
UCSD

Rowan Shi
Ryerson University

First Version: January 2017
This Version: March 2022
Most Recent Version

Abstract

Multinational enterprises (MNEs) pay high fixed costs to set up production sites around the globe. MNE's large plant networks lower marginal costs and require markups in pricing to recoup fixed cost outlays. Since plant networks often feature positive complementarities, productive MNEs grow large and charge high markups. Their size implies that MNE's pricing decisions drive aggregate markups in the economy. We propose a quantifiable model to study MNE's joint plant location and pricing decisions. We develop new computational methods to solve MNE's plant location problem with fixed costs and complementarities between plants and aggregate the resulting decisions over arbitrary distributions of firm heterogeneity. While our methods apply to a general class of multiple discrete choice problems, we use them to conduct general equilibrium counterfactuals on the impact of reductions in the costs of multinational activity on changes in aggregate markups.

*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. Code implementing the techniques introduced in this paper can be found at https://github.com/rowanxshi/CDCP.jl.

1 Introduction

Multinational enterprises operate global networks of plants that require fixed setup costs and allow them to lower their marginal cost of serving international markets. To recoup their fixed-cost investments, such firms charge a markup over their marginal cost. Complementarities between individual plants are essential for the profitability of additional plants and hence total fixed cost payment and markups. Understanding the price-setting behavior of such multinationals is crucial to understanding the evolution of aggregate markups since MNEs produce many products Americans use daily.[1] However, with complementarities and plant fixed costs, the number of potential plant location combinations a firm needs to consider grows exponentially in the number of locations available, and conventional solution methods fail. Moreover, aggregating optimal plant and pricing decisions across many heterogeneous firms is complicated because optimal decision sets can vary arbitrarily with firm type. As a result, the nexus of multinational production, firm-level, and aggregate markups has received little attention in the economics literature.

In this paper, we introduce a quantifiable model of multinational firms' joint plant location and pricing decisions. To formalize multinationals' joint plant location and pricing problem, we start with a model of multinationals' plant location decisions akin to Arkolakis et al. (2018) and add two features. First, we introduce fixed costs of opening foreign plants, implying that not all firms operate plants in all locations. Second, we choose a more general demand function that gives rise to variable markups but nests the constant markup, constant elasticity case as a limit (see, e.g., (Tintelnot, 2017)). In the model, firms choose a subset of countries to operate plants in and pay a plant-specific fixed cost. Individual plants can be complements or substitutes depending on parameters. As a result, an individual firm's plant location problem is to select a subset of countries to operate plants in when (a) each plant entails the payment of a fixed cost, and (b) the return to an individual plant depends on what other plants the firm operates. At the same time, the firm chooses the price charged for its product in each destination market, taking into account production costs that depend on its network of plants, trade costs, and local market conditions.

To solve the model, we develop a method to solve the firms' plant location problem with fixed costs and complementarities and to aggregate optimal plant choices across many heterogeneous firms. Our approach applies beyond the particular context of our model for general objective functions satisfying a set of easy-to-verify sufficient conditions. In particular, we define a class of multiple-discrete choice problems that we term "combinatorial discrete

[1] Bernard et al. (2009) report that multinationals accounted for almost 80% of U.S. imports and exports and employed 18% of the U.S. workforce in the year 2000

choice problems," or CDCPs for short. In CDCPs, an agent maximizes a return function $\pi$by choosing items from a discrete "choice set" $\mathcal{J}$, given an aggregate state $\mathbf{y}$. Formally:

Definition 1 (Combinatorial discrete choice problem)**.** A combinatorial discrete choice problem (CDCP) of an agent with type $\mathbf{z}$, given an aggregate state $\mathbf{y}$, is to find the optimal decision set $\mathcal{J}^\star$ such that

$$\mathcal{J}^\star = \arg\max_{\mathcal{J}\subseteq J} \pi(\mathcal{J};\mathbf{z},\mathbf{y})$$

where the "return function" π maps firm strategies, firm productivity, and aggregate state into profits. In our application of multinationals' pricing decisions, the agent is a firm and the set $\mathcal{J}$ a list of production locations.

Combinatorial discrete choice problems (cf. Definition 1) appear in many applications in economics. 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 between the individual items and with a pre-specified total number of choices (see, e.g., Hendel (1999)). In contrast, in CDCPs, the return to one item can depend on which others are chosen, and the size of the optimal decision set is not pre-specified.

We introduce an iterative solution method for CDCPs whose return function π satisfies a simple restriction that we refer to as single crossing differences in choices (SCD-C). The SCD-C assumption imposes structure on the complementarities between individual items in the return function.[2] There are two variants of the SCD-C condition, either of which works: SCD-C from above or SCD-C from below, corresponding to the case of negative and positive complementarities between items, respectively. The SCD-C assumption restricts the strength and direction of complementarities between different items in the agent's return function. For example, SCD-C from above implies that if an item adds value to a decision set, it continues to add value in any subset of the initial decision set. Supermodularity and submodularity of the return function are sufficient conditions for SCD-C from below and above, respectively.[3]

For CDCPs that satisfy the SCD-C condition, our method works by iteratively eliminating non-optimal decisions sets. Intuitively, by evaluating the return function at points of extreme

[2]The importance of single crossing differences in comparative statics analysis in mechanism design has been widely discussed (see Milgrom (2004)).

[3]Supermodularity and submodularity are of wide theoretical and practical use in economics in all contexts in which complementarities between choices are important. Supermodularity forces spillovers between decisions to always be positive, instead SCD-C from below allows for spillovers to be positive *or* negative, but restricts the strength of negative spillovers. Likewise submodularity implies that spillovers between decisions are always negative, while SCD-C from above only restricts them from being too positive.

complementarities, potential solutions can be discarded without considering them. For example, with SCD-C from above, items that increase the return function when all possible items are in the decision set must be part of the optimal decision set since they add value even when the negative complementarities are maximal. On the other hand, items that add no positive value to an otherwise empty decision set cannot add value to a decision set with other items that cause negative complementarities. Iteratively applying this logic shrinks the number of potential decisions sets without ever discarding the optimal decision set, $\mathcal{J}^\star$, and often isolates the optimal strategy.[4]

In our application, the net complementarities of locations in the firms' profit function are the result of three interacting forces. First, firms in our model serve locations without a plant from their plants in other countries via trade. This channel alone would make a firm's plants substitutes: a new plant means some incumbent plant no longer is the cheapest supplier for some destination markets. Second, firms with many plants already have low marginal costs, charge low prices, and sell large quantities in many destination markets. An additional cost reduction through an added plant is more valuable for such firms than firms that sell small amounts. This channel is active even without variable markups and, in isolation, would make plants complements. Third, variable markups amplify the previous channel since additional cost savings also allow firms to raise their markups, which is particularly valuable for large firms that produce and sell large quantities. In Arkolakis et al. (2018), only the first channel is active, plants are substitutes, and the profit function satisfies SCD-C from above. In Antras et al. (2017), the first and second channels are active. The authors assume the second overwhelms the first, making individual plants complements and the return function exhibit SCD-C from below. While our solution method works for both types of SCD-C, we assume that positive complementarities outweigh negative ones in our application.

To compute the general equilibrium of our model, we aggregate the location decisions of large numbers of heterogeneous firms that differ in their productivity. Classical discrete choice models assume that agents have idiosyncratic preferences over items (see McFadden (1973)). Aggregation of choices across agents occurs via distributional assumptions on these preferences. However, such an approach fails with CDCPs since it relies on the independence of irrelevant alternatives assumption that general complementarities between items, as in our setting, violate. Complementarities make aggregation particularly hard since they imply that firms with different productivity may have completely different optimal decision sets. While with return functions that satisfy SCDC from above, more productivity agent's choice sets

[4] 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.

always nest those of less productive ones, that is not the case with SCDC from below. We approach aggregation in two steps. First, we solve for the set-valued "policy function" that maps firm productivity to an optimal set of plant locations. Second, we integrate the policy function over arbitrary distributions of agent heterogeneity to compute general equilibrium objects.

We introduce a method to solve for such policy functions in a wide class of CDCPs. Our method does not require us to solve every agent's CDCP separately and does not rely on approximations. Our approach requires a second restriction on agents' return functions: single crossing differences in type (SCD-T).[5] SCD-T guarantees that agents form two contiguous groups in terms of their types for any potential decision set and any item: those that derive a positive marginal value from including an item in their optimal decision set and others that do not. Intuitively, SCD-T guarantees that similar agents have identical optimal decision sets so that the policy function only changes value at discrete points in the type space. We propose an iterative technique to solve for the policy function by identifying these points, and the optimal decision sets shared by the types between them.[6] In our applications, agents are firms, and their type is their core productivity shared by all production plants they control. Given this setup, firms' profit functions satisfy SCD-T so that our aggregation method is applicable.

A firm's optimal set of plant locations determines its marginal costs in each destination market and is hence an essential input into its price setting. Consumers in our application have demand functions as in Pollak (1971) which permit firms to charge variable markups. Since marginal costs are a crucial determinant of markups, multinational activity and the size of markups are naturally linked. In particular, we show how moving from N to $N+1$ plants leads to a discrete jump in a firm's markups in all destination markets. These results suggest that an expansion of multinational activity, e.g., due to a reduction in the cost of controlling foreign plants, has a direct and discrete impact on markups in all destination markets.

Our solution and aggregation methods are not just useful for counterfactual exercises but central to the calibration of our model. There is no longer a closed-form solution for aggregate trade flows in the presence of fixed costs of plant openings. To match our model with salient features of the international trade data, we employ a method-of-moments estimator that

[5]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. Return functions often exhibit single crossing differences as a consequence of natural economic assumptions.

[6]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.

requires recomputing the model for many different parameter guesses. Solving such a model many times would be computationally prohibitive without our solution and aggregation methods.

We use the calibrated model to study the role of multinational production networks for domestic markups. We highlight that plant-level fixed costs and plant complementarities that make a multinational's decision problem challenging to solve are also essential to the debate on markups. Recomputing our calibrated model without plant-fixed costs or with smaller positive complementarities between plants substantially lowers markups in the U.S. domestic market.

Literature Our paper contributes to two distinct literatures. First, we add 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.[7] These problems are quintessential CDCPs since the return to one location depends on which others are chosen. Most of the existing literature has either made these CDCPs trivial by abstracting from the fixed cost of adding a location (see Ramondo (2014), Ramondo and Rodríguez-Clare (2013), Arkolakis et al. (2018)) or solved them only for a small number of discrete locations so that it is feasible to evaluate the return to all possible combinations of location decisions separately (see Tintelnot (2017) and Zheng (2016)).[8] The difficulty of solving such problems has made it difficult to estimate the parameters of models involving them. Method of moment estimators require simulating agents' optimal decisions for different parameter values; when agents' problems are CDCPs this has often proved impossible due to difficulty of solving them, especially repeatedly. Several papers in economics have hence resorted to estimating the parameters of a model that features a CDCP using moment inequalities (see Morales et al. (2019), Holmes (2011)). Our method opens the way to estimate these models via simulation, and also to conduct counterfactuals once parameters have been estimated.

Jia (2008) presents a notable exception in introducing a "reduction method" that helps solve CDCPs with fixed costs when the return function is supermodular (see also Antras et al. (2017)). Our paper adds a new and unified way of solving CDCPs with positive or negative

[7]The classic Simple Plant Location Problem in operations research is an NP-hard problem that likewise fits the description in 1 (see Jakob and Pruzan (1983)).

[8]Without fixed costs, firms can trivially choose to include all locations in their choice set, even if some remain idle. Tintelnot (2017) uses 12 countries which allows computing the firm profits from any possible combination of plant locations of which there are 2^{12}. Zheng (2016) partitions the United States into small regional markets for which plant location decisions interact; decisions across these regional markets are independent, within them there a few enough locations to follow the "brute force" approach of Tintelnot (2017). The brute force approach becomes impossible with about 20 locations in most applications.

complementarities to this literature. Relative to Jia (2008), we present a generalized method applicable to return functions exhibiting a weaker form of positive spillovers, or, more importantly, negative spillovers. As explained above, such negative complementarities are an inherent feature of plant locations problems.[9] We also introduce a new method to aggregate combinatorial discrete choices across heterogeneous agents.

A second distinct contribution is our method for solving for the policy function mapping agent type to optimal decision, which together with the agent type distribution can be used to aggregate decisions. Random utility approaches (see McFadden (1973) or Eaton and Kortum (2002)) add random shocks to individual choices and assume them to be extreme value distributed which yields analytical expressions for the fraction of agents choosing each discrete alternative. However, these classical discrete choice models allow for one choice among mutually exclusive alternatives only (see Guadagni and Little (1983) and Train (1986)), so complementarities between different choices are not relevant by assumption.[10] As a result, standard discrete choice tools with random utilities are then not useful to study CDCPs.[11] Thus far, the literature has hence aggregated the solution to CDCPs solved by individual agents by discretizing the typespace and solving the CDCP only for a limited number of types, interpolating in-between them. 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.

We also contribute to the literature on mark-ups and multinationals. A small literature highlights the natural role of large, multinational firms could play for changing mark-ups due to correlation of markups and firm size (see, e.g., Antràs and Yeaple (2014)). However, most recent work on multinationals assumes away variable markups by focussing on constant elasticity of substitution demand systems (see, e.g., Tintelnot (2017), Arkolakis et al. (2018)). On the flipside, most international trade papers that do think about variable markups abstract from multi-plant location decisions to focus just on exporting versus FDI (see, e.g.,

[9]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.

[10]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). However, in these settings individual agents continue to choose just one alternative, but in the *aggregate* certain choices may be correlated, e.g., countries purchasing a lot from Germany may also purchase a lot from neighboring Austria, not for random reasons but due to a correlations of unobserved tastes or productivities.

[11]Hendel (1999) provides a notable exception by extending the random utility approach to multiple discrete choices, however, continues to abstract from interdependence between an agent's decisions.

Melitz and Ottaviano (2008)). Keller and Yeaple (2020) is an important exception: the paper provides direct empirical evidence for the role of markups in multinational companies, yet does not calibrate and solve a model with multinationals, markups, and complementarities between individual plants, which is the core contribution of our paper.[12]

2 A Theory of Multinationals, Plant Fixed Costs, and Variable Markups

The world economy consists of a set of countries $i = 1, ..., J$. Within each country there is a mass of heterogeneous firms, M_i, which are indexed by their efficiency z. Firms from a given country

choose a set of production locations $\mathcal{J} \subseteq J$ in which they establish plants. Firms incur a fixed cost $f_\ell > 0$ in order to set up a plant in location ℓ. The firm's revenue from setting up a plant in location ℓ may depend on what other locations are included in $\mathcal{J}$, rendering the decision problem *combinatorial* in nature. We now describe the firms problem more formally.

Production Technology Each plant of the firm produces the same unit continuum of firm-specific intermediate inputs. A firm headquartered in country i that has productivity z in its domestic plant can produce any of its varieties at a plant in country ℓ with a fraction $1/\gamma_{i\ell} < 1$ of its domestic productivity. All intermediate production is labor-only and a firm headquartered in i with productivity$z$ has the following unit cost to produce any of its varieties at a plant in location ℓ:

$$c_{il} = \frac{\gamma_{i\ell} w_\ell}{z},$$

where w_ℓ is the cost of a unit of labor in country ℓ.

Final consumers in each country demand the firm's *final* good which is a CES composite of its intermediate inputs:

$$q_\omega = \left[\int_\nu q(\nu)^{\frac{\eta-1}{\eta}} d\nu \right]^{\frac{\eta}{\eta-1}}$$

The firm assembles its final composite good in each destination market separately by import-

[12]Other papers model instead the decision of firms with which countries to establish sourcing relationships (see Antras et al. (2017)), or outsourcing strategies in the presence of industry complementarities (see Jiang and Tyazhelnikov (2020)). In this setting, adding a new supplier reduces production costs, but reduces sales from each other supplier, lowering their marginal value, so that decisions are again interdependent. In other branches of economics researchers study the optimal evolution of networks, where one new link may beget others, or the optimal ATM or hospital network of locations, both can be formulated as CDCPs.

ing the intermediate inputs from its various plants. For each intermediate input, it chooses the plant that can deliver it at the lowest marginal cost. Exporting from a plant in ℓ to a final destination market n incurs iceberg trade costs and a product-plant specific cost as in Tintelnot (2017). In particular, the unit cost of delivering variety $\nu$from a plant in ℓ to destination $n$for a firm from i with productivity z is given by

$$c_{iln}(\nu) = \frac{c_{il}\tau_{ln}}{\varphi_\ell(\nu)}.$$

We denote the whole vector of plant-ℓ-specific costs by $\boldsymbol{\varphi}_\ell = \{\varphi_\ell(\nu)\}$. We assume that $\boldsymbol{\varphi}_\ell$ is unknown to the firm when choosing its plant locations, but known when choosing the optimal plant ℓ to serve market n. The firm maximizes its *expected return* when choosing its plant locations $\mathcal{J}$ prior to the realization of $\boldsymbol{\varphi}_\ell$ based on its expected realization.

For every destination market n, the firm chooses the best plant location ℓ to produce each variety ν,

$$x_{in} \equiv \min_{j\in\mathcal{J}} \frac{\gamma_{i\ell}w_\ell\tau_{\ell n}}{z\varphi_\ell}.$$

We let $G_{in}(x;z)$ describe the distribution of x_{in}, conditional on the firm's productivity z. Then, the marginal cost of producing one unit of firm i's bundle for market n is

$$c_{in}(\omega) = \left[\int_\varphi x^{1-\eta}\mathrm{d}G_{in}(x;z)\right]^{\frac{1}{1-\eta}}.$$

If we assume that that the firm draws the vector $\boldsymbol{\varphi}_\ell$ for each plant and variety is drawn independently from a Frechet distribution with scale parameter T_ℓ we obtain the following expression for firm i's expected unit cost of delivering a unit of its final good to destination n given its productivity and a potential plant network $\mathcal{J}$:

$$c_{in}(\mathcal{J};z) = \tilde{\Gamma}\frac{1}{z}\left[\sum_{\ell\in\mathcal{J}}\left(\gamma_{i\ell}w_\ell\tau_{\ell n}/T_\ell\right)^{-\theta}\right]^{-\frac{1}{\theta}}$$

where $\tilde{\Gamma}$ is a constant of integration.[13] The unit cost increase in the wages, trade costs and productivity loss associated with each plant l and decrease in firm efficiency z. Marginal costs are a function of firm type and production locations $\mathcal{J}$. We define

$$\Theta_{in}(\mathcal{J}) \equiv \sum_{\ell\in\mathcal{J}}\left(\gamma_{i\ell}w_\ell\tau_{\ell n}/T_\ell\right)^{-\theta} \equiv \sum_{\ell\in\mathcal{J}}\xi_{jn}$$

[13] Using instead a plant-specific multivariable Pareto distribution as in Arkolakis et al. (2018) yields a similarly nice expression.

as the *production potential* of a particular set of plant locations $\mathcal{J}$ for the destination market n. Note that individiual firm efficiency just shifts this production potential by a factor $1/z$, so that the relative production potentials of different firms from i in destination market n is just the inverse of their relative efficiencies.

Demand We follow Arkolakis et al. (2019) and specify a generalized demand function faced by the firm in each market n given by

$$q_n(z) = Q_n D(p_n(z)/P_n)$$

where Q_n is an aggregate demand shifter and P_n is an aggregate price shifter both of which are matket aggregates determined in general equilibrium. The erm $p_n(z)$ is the price set by a firm of productivity z in destination market n.[14] We denote the vector of aggregates faced by the firm as $\mathbf{y} = [\mathbf{Q}, \mathbf{P}, \mathbf{w}]$, and without risk of confusion we include in $\mathbf{y}$ information on the specific aggregates relevant for each market.

Profit maximization implies that firms set the price for their final good as a markup over their destination specific marginal cost c_{in} as follows:

$$p_{in}(z) = \frac{\varepsilon_D(p(c_{in}(\mathcal{J};z)))}{\varepsilon_D(p(c_{in}(\mathcal{J};z))) - 1} c_{in}(\mathcal{J};z) \equiv \mu(c_{in}(\mathcal{J};z))\, c_{in}(\mathcal{J};z) \tag{1}$$

where ε_D is the elasticity of the demand D and μ is the optimal markup as a function of marginal cost. Equation 1 is important as it shows that the markup is a direct function of the multinational plant location choice of the firm: plant location choices and markups in each country are co-determined. Importanlty, a firms markups in market n depend on its full plant location network. So changes in a multinational firm's domestic markups may be the result of an expansion of its international plant network.

Profits and Firm Optimization Finally, given a set of production locations $\mathcal{J}$, the firms variable profits in market n are

$$\pi_n(c_{in}(\mathcal{J};z), \mathbf{y}) \equiv (\mu(c_{in}(\mathcal{J};z)) - 1)\, QD(p(c_{in}(\mathcal{J};z))/P)\, c_{in}(\mathcal{J};z).$$

[14] 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)).

The firm's total profits across all markets are given by the function

$$\pi(\mathcal{J};\mathbf{z},\mathbf{y}) = \sum_{n} \pi_n \left(c_{in}\left(\mathcal{J};z\right),\mathbf{y}\right) - \sum_{\ell\in\mathcal{J}} f_\ell.$$

The firm then chooses $\mathcal{J}$ so as to maximize its total profits making its location choice problem a combinatorial discrete choice problem as in Definition 1.[15] We devote the next section to the description of a new method to solve the firm's location choice problem that is applicable beyond the context of our particular model as along as a set of conditions on the profit function are met. In Section 4, we return to the model introduced here, validate that it satisfies the conditions for our solution method to be applicable. We then close the model, calibrate it, and provide a set of general equilibrium counterfactuals.

3 Solving and Aggregating CDCPs

Instead of focussing on the specific problem of the firm outlined in the previous section, we develop a general method for CDCPs as in Definition 1.

We consider an economy inhabited by a set of heterogeneous agents indexed by their type $\mathbf{z} \in \mathbf{Z} \subseteq \mathbb{R}^N$. The "aggregate state" of the economy is $\mathbf{y} \in \mathbf{Y} \subseteq \mathbb{R}^M$. Agents choose a subset $\mathcal{J}$ of items from a finite discrete set J in order to maximize a return function of the form

$$\pi(\mathcal{J};\mathbf{z},\mathbf{y}) : \mathscr{P}(J) \times \mathbf{Z} \times \mathbf{Y} \to \mathbb{R}, \tag{2}$$

where the power set $\mathscr{P}(J) = \{\mathcal{J} \mid \mathcal{J} \subseteq J\}$ is the collection of all possible subsets of J. We refer to $\mathcal{J}$ as the firm's "decision set," to J as their "choice set," and to $\mathscr{P}(J)$ as their "choice space."[16]

To formalize the interdependence of decisions, we introduce a marginal value operator that encodes the additional benefit agents derive from element j's inclusion in the decision set $\mathcal{J}$.

[15] Note that the fixed costs can be firm-specific in which case they would form part of the (multidimensional) firm type $\mathbf{z}$.

[16] The formulation in equation 2 is very general. It could be an individual's utility function, and J the collection of items chosen in the supermarket. In many applications other continuous variables are chosen conditional on the choice of a discrete set $\mathcal{J}$ of items, e.g., conditional on having a plant in location A how much should this plant produce? In such situations, the return function can be expressed as in equation 2 after "maximizing out" the continuous choice variable. Dynamic combinatorial discrete choice problems can be written as in equation 2. Suppose each period a firm can choose in which additional locations to open plants. Equation 2 would then refer to the static problem of choosing the optimal additional plants in a given period, given the plants already present (which have to be removed from the choice set J) and given the future paths of plant decisions that would follow any decision today.

Definition 2 (Marginal value operator)**.** For an item $j \in J$, decision set $\mathcal{J}$, firm type $\mathbf{z}$, and aggregate vector $\mathbf{y}$, the marginal value operator D_j on the return function $\pi(\mathcal{J}; \mathbf{z}, \mathbf{y})$ is defined as

$$D_j \pi(\mathcal{J}; \mathbf{z}, \mathbf{y}) \equiv \pi(\mathcal{J} \cup \{j\}; \mathbf{z}, \mathbf{y}) - \pi(\mathcal{J} \setminus \{j\}; \mathbf{z}, \mathbf{y}) . \tag{3}$$

Decisions are *interdependent* as long as the marginal value $D_j \pi(\mathcal{J}; \mathbf{z}, \mathbf{y})$ of item j depends on the overall choice set $\mathcal{J}$. If the marginal value of any item j is the same across all decision sets $\mathcal{J}$, decisions are independent and the agent can simply consider each item in isolation.

Our multinational plant location problem is an example of a classical CDCP one where the firm's plants compete to serve demand points in each location (see Balinski (1965), Owen and Daskin (1998), Ramondo (2014), Ramondo and Rodríguez-Clare (2013), Arkolakis et al. (2018), Tintelnot (2017), Zheng (2016), Jia (2008)). In these problems adding an additional plant in a given location may save on trade costs to serve the location's consumers, but takes away from the sales of other plants that served this location via trade, making plant location decisions inherently interdependent.

In many of these applications, the economy is populated by a large number of agents whose return functions may differ. For such settings, we additionally 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, conditional on the aggregate state.

Definition 3 (Policy function)**.** Consider a combinatorial discrete choice problem (CDCP) confronted by agents of heterogeneous types $\mathbf{z}$ and given an aggregate state $\mathbf{y}$. The policy function mapping agent type to an optimal decision set, conditional on the aggregate state $\mathbf{y}$, is defined as

$$\mathcal{J}^{\star}(\cdot; \mathbf{y}) : \mathbf{Z} \to \mathscr{P}(J)$$

where the return function $\pi(\mathcal{J}; \mathbf{z}, \mathbf{y})$ is defined in equation 2.

The policy function is useful for the aggregation of decisions across firms. Together with an arbitrary distribution of heterogeneity, $f(z)$, the policy funcition can be used to compute aggregate quantities and prices in the economy. The classicdiscrete choice literature insisted on extreme-value distributions to aggregate discrete choices across heterogeneous agents (see McFadden (1973)) but its approach does not generalize to settings such as ours where agents can choose an arbitrary subset of the available choices instead of just choosing one, utility-maximizing alternative. Our numerical approach is that its designed for situations in which individual agents make interacting choices and works with arbitrary distributions of agent heterogeneity. It is thus ideally suited to handle the combinatorial discrete choice of the multinational firm that we study.

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

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 2. For a given type $\mathbf{z}$ and aggregate vector $\mathbf{y}$,

The return function obeys single crossing differences in choices from above if, for all elements $j \in J$ and decision sets $\mathcal{J}_1 \subset \mathcal{J}_2 \subseteq J$,

$$D_j\pi\left(\mathcal{J}_2; \mathbf{z}, \mathbf{y}\right) \geq 0 \qquad \Rightarrow \qquad D_j\pi\left(\mathcal{J}_1; \mathbf{z}, \mathbf{y}\right) \geq 0\,.$$

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

$$D_j\pi\left(\mathcal{J}_1; \mathbf{z}, \mathbf{y}\right) \geq 0 \qquad \Rightarrow \qquad D_j\pi\left(\mathcal{J}_2; \mathbf{z}, \mathbf{y}\right) \geq 0\,.$$

These restrictions are intuitive. For return functions satisfying SCD-C from above, if the marginal value of including an additional element j in a given decision set is positive, it remains so as elements are removed from the decision set. Similarly, for return functions staifsying SCD-C from below, if the marginal value of including an additional element j 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 j, for all $j \in J$, to be monotone in its first argument. In particular, given any two sets $\mathcal{J}_1 \subseteq \mathcal{J}_2$, if

$$D_j\pi(\mathcal{J}_1; \mathbf{z}, \mathbf{y}) \geq D_j\pi(\mathcal{J}_2; \mathbf{z}, \mathbf{y}) \quad \forall j \in J,$$

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.[17]

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, $\mathbf{z}$, and aggregate state, $\mathbf{y}$, are held fixed in this section, we omit them for notational brevity.

Consider the choice set J of the CDCP defined in equation (2). We introduce an associated pair of sets $[\underline{\mathcal{J}}, \overline{\mathcal{J}}]$, which we call the "bounding sets." We use these sets to keep track of items from J that are certain to be included in or excluded from the optimal decision set $\mathcal{J}^\star$. The "subset" $\underline{\mathcal{J}}$ *includes* all items in $\mathcal{J}$ we know to *be* in the optimal decision set. The "superset" $\overline{\mathcal{J}}$ *excludes* all items in $\mathcal{J}$ we know to *not be* in the optimal decision set. The set difference of $\overline{\mathcal{J}}$ and $\underline{\mathcal{J}}$, denoted $\overline{\mathcal{J}} \setminus \underline{\mathcal{J}}$, 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 $\underline{\mathcal{J}} = \emptyset$ and $\overline{\mathcal{J}} = J$, so that $\overline{\mathcal{J}} \setminus \underline{\mathcal{J}} = J$, that is, all items in J are undetermined.

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

[17]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.

step, we introduce an auxiliary mapping

$$\Omega(\mathcal{J}) \equiv \{j \in J \mid D_j \pi(\mathcal{J}) > 0\}$$

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

Definition 5 (Squeezing step)**.** Consider a CDCP and its associated bounding sets $[\underline{\mathcal{J}}^{(k)}, \overline{\mathcal{J}}^{(k)}]$. The mapping S^a is such that

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

the mapping S^b is such that

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

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, $\underline{\mathcal{J}}$, while removing elements from the superset, $\overline{\mathcal{J}}$, 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 (2).*

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

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

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.[18] In particular, the expression $\underline{\mathcal{J}}^{(k)} \subseteq \underline{\mathcal{J}}^{(k+1)}$ implies that (weakly) more

[18]The squeezing step is designed to recover $\mathcal{J}^\star$ so that all items j for which the agent is indifferent are

items are *included* in the subset — and hence known to be in the optimal decision set — after applying the squeezing step. Similarly, the expression $\overline{\mathcal{J}}^{(k+1)} \subseteq \overline{\mathcal{J}}^{(k)}$ 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 $\underline{\mathcal{J}}^{(k+1)} \subseteq \mathcal{J}^{\star} \subseteq \overline{\mathcal{J}}^{(k+1)}$.

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 $[\underline{\mathcal{J}}^{(K)}, \overline{\mathcal{J}}^{(K)}]$ the resulting bounding sets. In the Appendix, we establish that K is never larger than the cardinality of the choice set, $|J|$.

Consider the bounding sets resulting from applying mapping S^a or S^b until convergence. If the converged pair of bounding sets is identical such that $\underline{\mathcal{J}}^{(K)} = \overline{\mathcal{J}}^{(K)}$, Theorem 1 implies that $\mathcal{J}^{\star} = \underline{\mathcal{J}}^{(K)} = \overline{\mathcal{J}}^{(K)}$ so that we have identified the optimal decision set solving the CDCP. Notice that a limitation of our approach is that the return function has to satisfy the same type of SCD-C (i.e., either below or above) over the entire choice space.[19]

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 choice space. Instead, we introduce a "branching procedure", described and formally characterized in Appendix 8, 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.

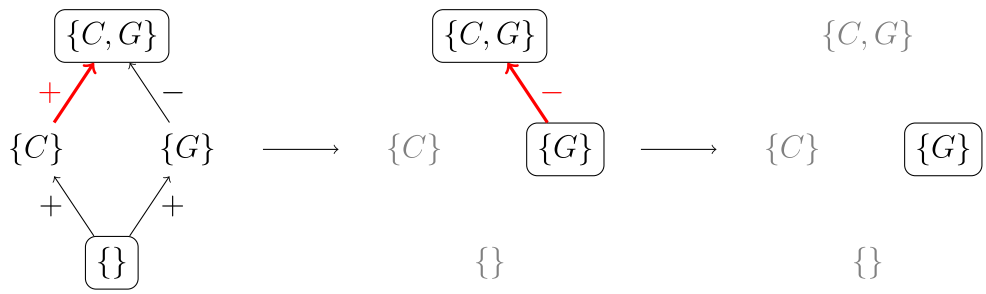
A Simple Example To elucidate our solution concept, we consider a simple example of our multinationals framework where there are only two available countries, Germany (G) and Canada (C). The firm's choice set is $J = \{G, C\}$ and its choice space is $\mathscr{P}(J) = \{\{\}, \{G\}, \{C\}, \{G, C\}\}$, which contains all potential decision vectors of the firm.

Consider the marginal value of building a plant in Germany, $\mathcal{J} = \{G\}$, rather than no plant at all, $\mathcal{J} = \{\}$. What we learn from this marginal value depends on its sign and the type of complementarities between individual plants in the firm's return function. For illustration, suppose the marginal value is negative and the return function features negative

excluded from the optimal strategy. If these items should be included, they are easily identified as those j for which $D_j\pi(\mathcal{J}^{\star}) = 0$.

[19]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.

Figure 1: 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.

complementarities between plants. But then the negative marginal value implies that the optimal plant strategy does not involve a plant in Germany: having just a plant in Germany yields a negative return, adding another plant in Canada would further lower the marginal return to the German plant due to the negative complementarities.

Consider again the firm choosing optimal plant locations when its choice set is Germany and Canada. Suppose the firm's return function satisfies SCD-C from above. Figure 1 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 right most panel. The initial bounding pair is $[\{\}, \{G, C\}]$; $\mathcal{J}^\star \in [\underline{\mathcal{J}}, \overline{\mathcal{J}}]$, as required. Beginning with the subset, $\underline{\mathcal{J}} = \{\}$, 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, $\overline{\mathcal{J}} = \{G, C\}$. 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 1 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 $\mathcal{J}^\star = \{G\}$.

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

$$\Lambda_j(\mathcal{J}) = \{\mathbf{z} \in \mathbf{Z} \mid D_j(\mathcal{J}; \mathbf{z}) > 0\}$$

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 j and decision sets $\mathcal{J}$, $\Lambda_j(\mathcal{J})$ and its complement $\Lambda_j(\mathcal{J})^c$ are both connected sets.

The two contiguous sets $\Lambda_j(\mathcal{J})$ and $\Lambda_j(\mathcal{J})^c$ divide the typespace $\mathbf{Z}$ into types which receive positive marginal value and types which receive negative marginal value from j's inclusion in $\mathcal{J}$.[20]

Intuitively, the SCD-T restrictions implies that if the addition of item j to choice set $\mathcal{J}$ 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 typespace.

If type heterogeneity is one-dimensional, 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 $j \in J$, decision sets $\mathcal{J}$, and aggregate vectors $\mathbf{y}$,[21]

$$D_j\pi(\mathcal{J}; z_1, \mathbf{y}) \geq 0 \qquad \Rightarrow \qquad D_j\pi(\mathcal{J}; z_2, \mathbf{y}) \geq 0\,.$$

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 submodularity when decreasing).

[20] The SCD-T property is therefore implied by the supermodularity property introduced by Costinot (2009) (in its log form). Notice that our analysis is focused on partially ordered sets (lattices) not necessarily on totally ordered sets. We discuss below preliminary results related to the analysis in Costinot (2009) when the policy function obeys a *nesting* structure.

[21] 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.

A sufficient condition for SCD-T follows from super- and sub-modularity between type and decision set. In particular, fix an item j, set $\mathcal{J}$, and component i of the multi-dimensional type vector. Holding all other components of the type vector fixed, one can check if the selected component i exhibits either supermodularity or submodularity with the decision set. If it does for every possible item j, set $\mathcal{J}$, and component i, then the return function exhibits SCD-T.[22]

In what follows, we use the SCD-T restriction to solve for the *policy function*, $\mathcal{J}^\star(\cdot)$, that maps agents' types to their optimal decision sets. We re-introduce the $\mathbf{z}$ indexing to indicate an agent's type, but continue to omit the aggregate state $\mathbf{y}$. 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.

Intuitively, 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 single-dimensional typespace obeying these restrictions. In the figure, agents with types between z_1 and z_2 have the same optimal decision set $\mathcal{J}_1^\star$, while types between z_2 and z_3 instead optimally choose $\mathcal{J}_2^\star$, 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.

In the special case of single dimensional type heterogeneity with a 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 $\mathcal{J}^\star(z_1) \subseteq \mathcal{J}^\star(z_2)$.[23] In the appendix, we generalize this nesting result to a multidimensional typespace 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 item than less productive agents, and vice versa. The resulting optimal policy function is a complicated object that is difficult to theoretically characterize. This challenge

[22] This sufficient condition still remains relatively general. For example, it allows for a given component i to be supermodular with item and decision set $(j, \mathcal{J})$, but submodular with different pair $(j', \mathcal{J})$. Likewise, it allows for a given component i to be supermodular with a pair $(j, \mathcal{J})$, while another component i' is submodular with the same pair.

[23] 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 typespace 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

$$\underline{z} \;\overset{\mathcal{J}_1^\star}{\rule{3cm}{0.4pt}}\; z_1 \;\overset{\mathcal{J}_2^\star}{\rule{1.5cm}{0.4pt}}\; z_2 \;\overset{\mathcal{J}_3^\star}{\rule{1cm}{0.4pt}}\; z_3 \;\overset{\mathcal{J}_4^\star}{\rule{2cm}{0.4pt}}\; \overline{z}$$

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 $n = 1, 2, 3$.

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.

As illustrated in Figure 2, solving for the policy function requires *both* finding 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

In the single agent CDCP in Section 3.2, we introduced the notion of bounding sets, $[\underline{\mathcal{J}}, \overline{\mathcal{J}}]$ associated with a CDCP. The set $\underline{\mathcal{J}}$ *includes* all items in $\mathcal{J}$ we know to *be* in the optimal decision set, while the set $\overline{\mathcal{J}}$ *excludes* all items in $\mathcal{J}$ we know to *not be* in the optimal decision set. As a result, $\underline{\mathcal{J}} \subseteq \mathcal{J}^\star \subseteq \overline{\mathcal{J}}$. With heterogeneous agents, we extend the notion of bounding sets to set-valued functions over the typespace, $\underline{\mathcal{J}}(\cdot)$ and $\overline{\mathcal{J}}(\cdot)$. These "boundary set functions" are such that $\underline{\mathcal{J}}(\mathbf{z}) \subseteq \mathcal{J}^\star(\mathbf{z}) \subseteq \overline{\mathcal{J}}(\mathbf{z})$ 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 typespace Z such that for all $\mathbf{z} \in Z$ the current boundary set functions yield a constant value. We collect all information on a region Z relevant to the squeezing step in a 4-ple, $[(\underline{\mathcal{J}}, \overline{\mathcal{J}}, M), Z]$, where $\underline{\mathcal{J}}$ and $\overline{\mathcal{J}}$ are the values of the current bounding set functions over the interval Z. The set M is an "auxiliary" set which 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 typespace 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-ple creates up to three new 4-ples, 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-ples from an initial 4-ple at each application. The eventual output is a collection of 4-ples each with associated boundary sets. As with the simple squeezing procedure, ideally the boundary sets of each 4-ples coincide with one another in which case they also coincide with the optimal strategy for all types in the associated subregion of the typespace Z.

A natural initiation for applying the generalized squeezing procedure to a CDCP is to set the boundary sets to reflect that all items are undetermined, i.e., $\underline{\mathcal{J}}(\mathbf{z}) = \emptyset, \overline{\mathcal{J}}(\mathbf{z}) = J, \forall z \in Z$, and the auxiliary set to reflect that no item has been tried, i.e., $M = \emptyset$. Correspondingly, the initial 4-ple contains the entire typespace, i.e., $Z = \mathbf{Z}$.

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

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

The mapping S^a is defined as

$$S^a([(\underline{\mathcal{J}}, \overline{\mathcal{J}}, M), Z]) \equiv \{[(\underline{\mathcal{J}} \cup \{j\}, \overline{\mathcal{J}}, \emptyset), \Lambda_j(\overline{\mathcal{J}})], [(\underline{\mathcal{J}}, \overline{\mathcal{J}} \setminus \{j\}, \emptyset), \Lambda_j(\underline{\mathcal{J}})^c)],$$

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

The mapping S^b is defined as

$$S^b([(\underline{\mathcal{J}}, \overline{\mathcal{J}}, M), Z]) \equiv \{[(\underline{\mathcal{J}} \cup \{j\}, \overline{\mathcal{J}}, \emptyset), \Lambda_j(\underline{\mathcal{J}})], [(\underline{\mathcal{J}}, \overline{\mathcal{J}} \setminus \{j\}, \emptyset), \Lambda_j(\overline{\mathcal{J}})^c)],$$

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

We use an example, to show how to use the generalized squeezing step. Consider a CDCP with a single dimension of heterogeneity and with a return function that satisfies SCD-T and SCD-C from above. In the initial 4-ple Z is set to the entire typespace $\mathbf{Z}$, $\underline{\mathcal{J}}(\cdot) = \emptyset, \overline{\mathcal{J}}(\cdot) = J, \forall z \in Z$, and $M = \emptyset$. To apply the generalized squeezing step S^a to the corresponding 4-ple $[(\underline{\mathcal{J}}, \overline{\mathcal{J}}, M), Z]$, we first choose an undetermined item $j \in \overline{\mathcal{J}} \setminus \underline{\mathcal{J}}$. We can identify the "cutoff" agent types $z^{\text{in}}, z^{\text{out}} \in Z$ which are exactly indifferent between including j in $\overline{\mathcal{J}}$ and

$\underline{\mathcal{J}}$ respectively:

$$0 = D_j\pi(\overline{\mathcal{J}}; z^{\text{in}}) \qquad\qquad 0 = D_j\pi(\underline{\mathcal{J}}; z^{\text{out}}) .$$

These cutoff types divide the original region Z into *up to* three subregions.[24]

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

$$0 = D_j\pi(\underline{\mathcal{J}}; z^{\text{out}}) \quad \Rightarrow \quad 0 \geq D_j\pi(\mathcal{J}^\star(z); z^{\text{out}}) \quad \Rightarrow \quad 0 \geq D_j\pi(\mathcal{J}^\star(z); z) .$$

The first inequality follows from SCD-C, since $\underline{\mathcal{J}} \subseteq \mathcal{J}^\star(z)$ 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 $z > z^{\text{in}}$,

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

Again, the first inequality follows from SCD-C, since $\mathcal{J}^\star(z) \subseteq \overline{\mathcal{J}}$ 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-ples. After dividing Z into three new subregions according to the two cutoffs z^{in} and z^{out}, we update each subregion's bounding and auxiliary sets with the new information. For all types in the right subregion, j is optimally included, so the subset $\underline{\mathcal{J}}$ includes j. For all types in the left subregion, j is optimally excluded, so the superset $\overline{\mathcal{J}}$ excludes j. For all types in the middle subregion, we cannot conclude that j is either optimally included or excluded. Its bounding sets, $\underline{\mathcal{J}}$ and $\overline{\mathcal{J}}$, are the same as those of its "parent" region Z. For this intermediate region, we instead add j to M to encode the information that, given the current bounding sets $[\underline{\mathcal{J}}, \overline{\mathcal{J}}]$ on Z, j remains undetermined. Consequently, the set M must be "reset" to the empty set whenever either bounding set, $\underline{\mathcal{J}}$ or $\overline{\mathcal{J}}$, is non-trivially updated since this information changes the potential marginal value of the items in M.[25] We have now created three new subregions from the original subregion Z, each with at least one of the original $\underline{\mathcal{J}}$, $\overline{\mathcal{J}}$, or M updated.[26] Note that it could be that $z^{\text{out}} \notin Z$ of $z^{\text{in}} \notin Z$ or $z^{\text{out}}, z^{\text{in}} \notin Z$ in which case the generalized

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

[25] Observe that the item j must therefore be chosen from $\overline{\mathcal{J}} \setminus (M \cup \underline{\mathcal{J}})$. The reason is that, for the current bounding sets, all items in M have already been considered with no progress made.

[26] Returning to the definition of **??**, we can verify the above example in single dimensional typespace 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.

squeezing step creates less than three new 4-ples.

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

Theorem 2. *Consider a CDCP as defined in equation (2) for agents on a typespace* $\mathbf{z}$*, and associated 4-ple* $[(\underline{\mathcal{J}}_0, \overline{\mathcal{J}}_0, M), Z]$ *for which* $M \subseteq (\overline{\mathcal{J}}_0 \setminus \underline{\mathcal{J}}_0)$ *and* $\overline{\mathcal{J}}_0 \subseteq \mathcal{J}^\star(\mathbf{z}) \subseteq \underline{\mathcal{J}}_0$ *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,* $\underline{\mathcal{J}}_0 \subseteq \underline{\mathcal{J}}(\mathbf{z}) \subseteq \mathcal{J}^\star(\mathbf{z}) \subseteq \overline{\mathcal{J}}(\mathbf{z}) \subseteq \overline{\mathcal{J}}_0$ *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,* $\underline{\mathcal{J}}_0 \subseteq \underline{\mathcal{J}}(\mathbf{z}) \subseteq \mathcal{J}^\star(\mathbf{z}) \subseteq \overline{\mathcal{J}}(\mathbf{z}) \subseteq \overline{\mathcal{J}}_0$ *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 $\overline{\mathcal{J}} = M \cup \underline{\mathcal{J}}$ on each subregion of the typespace. The lower portion of Figure **??** visualizes a possible "tree" of subregions that emerges from applying the generalized squeezing step repeatedly to each new 4-ple.

Once $\overline{\mathcal{J}} = M \cup \underline{\mathcal{J}}$ for each subregion, there remain no undetermined elements j on which to make progress. The squeezing procedure has converged globally when it converges on all 4-ples separately. We denote the operator that encodes the recursive application of the generalized squeezing step until global convergence by $S^{a(K)}$ and $S^{b(K)}$. Given Theorem 2, the generalized squeezing procedure delivers bounding set functions $\underline{\mathcal{J}}(\cdot)$ and $\overline{\mathcal{J}}(\cdot)$ such that $\underline{\mathcal{J}}(\mathbf{z}) \subseteq \mathcal{J}^\star(\mathbf{z}) \subseteq \overline{\mathcal{J}}(\mathbf{z})$ for all $\mathbf{z} \in \mathbf{Z}$.

For subregions where $\underline{\mathcal{J}}(\mathbf{z}) \subset \overline{\mathcal{J}}(\mathbf{z})$ for some type $\mathbf{z}$ after global convergence of the generalized squeezing procedure, we provide a generalized branching procedure in the Appendix.

Back to the Simple Example We now revisit the simple example from before, but with several firms whose return functions satisfies SCD-C from above and SCD-T. Each row of Figure 3 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.

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.

The middle row of Figure 3 shows the typespace with updated bounding pairs. For the middle interval, G enters the auxiliary set M since no decision could be reached yet; we encode this update by representing Germany in blue.

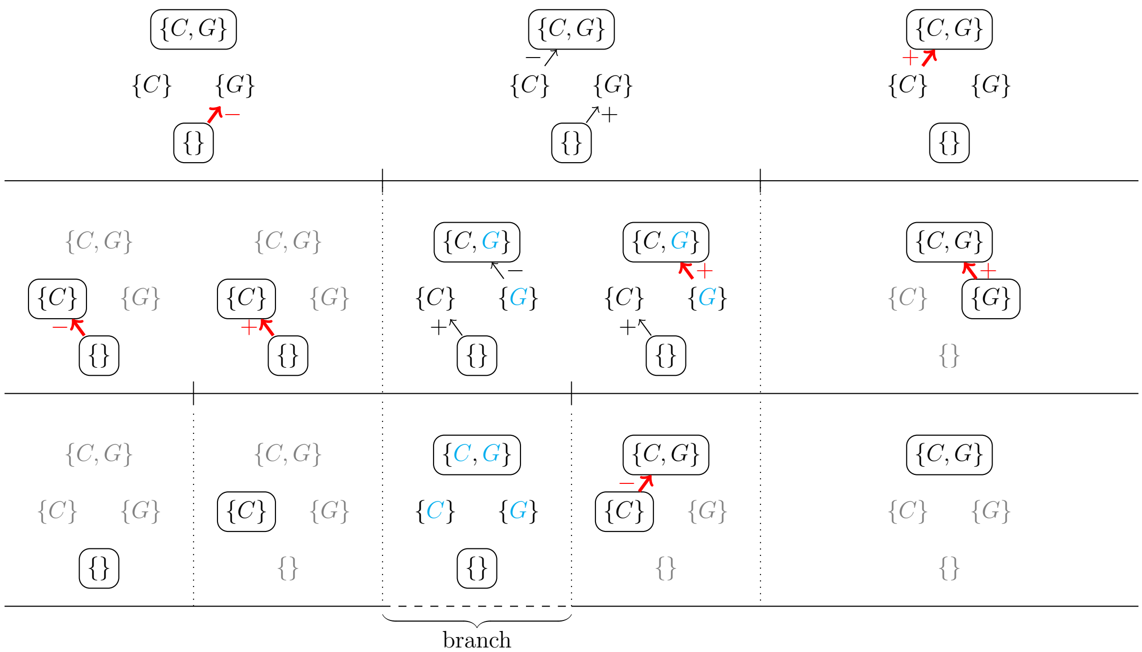
In the leftmost interval, we identify the cutoff productivity for which opening a plant in Canada has positive marginal value. In the rightmost interval, all types receive positive marginal value from opening a German plant. In the middle interval, we identify a cutoff above which firms receive positive marginal benefit from a Canadian plant when a German plant exists.

The third row of Figure 3 reflects updates from the second line. 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 Closing and Solving the Model of Multinationals' Pricing and Location Decisions

In this section, we provide sufficient conditions for SCD-C and SCD-T to hold in our multinational plant location problem with variable markups. We then specify a demand function that satisfies this conditions and consider the general equilibrium of our setup.

Figure 3: An Example of the Generalized Squeezing Procedure



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

4.1 General Conditions that satisfy SCD-C and SCD-T

For the multinational model introduced in the previous section, we can validate that the SCDC proprty holds for the firm's return function. We drop the n indexing where umambiguous. A location's marginal benefit given decision set $\mathcal{J}$ can be written as

$$\begin{aligned}\pi(c(\mathcal{J}\cup\ell))-\pi(c(\mathcal{J})) &= \sum_n\left[\tilde{\pi}_n\left(\xi_{\ell n}+\sum_{j\in\mathcal{J}}\xi_{jn}\right)-\tilde{\pi}_n\left(\sum_{j\in\mathcal{J}}\xi_{jn}\right)\right]-f_\ell \\ &\equiv \sum_n\left[\tilde{\pi}_n\left(\Theta_{in}(\mathcal{J}\cup\ell)\right)-\tilde{\pi}_n\left(\Theta_{in}(\mathcal{J})\right)\right]-f_\ell\end{aligned}$$

where $\tilde{\pi}_n(\Theta)\equiv\pi_n(\tilde{\Gamma}\Theta^k/z)$ are variable profits in market n as a function of production potential. We establish in Appendix 7 that signing the second derivative of destination market-specific return function with respect to the production potential ($\tilde{\pi}_\mathrm{n}''\equiv\mathrm{d}^2\tilde{\pi}_n/\mathrm{d}\Theta^2$) for each n is sufficient to establish SCD-C. In particular, $\tilde{\pi}_n''\geq 0$ implies positive complementarities and is thus sufficient to guarantee SCD-C from below (and similarly $\tilde{\pi}_n''\leq 0$ for SCD-C from above). The second derivative of the destination-specific profit function with respect to the production is given by:

$$\frac{\mathrm{d}^2\tilde{\pi}}{\mathrm{d}\Theta^2} = -\pi'\left(\frac{\tilde{\Gamma}\Theta^{-\frac{1}{\theta}}}{z}\right)\frac{\tilde{\Gamma}\Theta^{-\frac{1}{\theta}-2}}{z}\left(-\frac{1}{\theta}\right)^2\left[\varepsilon_{\pi'} - (1+\theta)\right] \tag{4}$$

where the elasticity of the derivative of the profit function is

$$\varepsilon_{\pi'} = -\frac{\pi''(c)}{\pi'(c)}c.$$

The sign of $\frac{\mathrm{d}^2\tilde{\pi}}{\mathrm{d}\Theta^2}$ is determined by the component in square brackets since the term pre-multiplying them is positive ($\pi' \leq 0$). So the firms profit function in our model satisfies SCD-C from below if the term in square brackets is positive, i.e., if $\varepsilon_{\pi'}$ exceeds$1+\theta$, and SCDC-C from above if the term is negative.

Lastly, we establish the conditions under which SCD-T holds in our model. Notice that in our framework maginal costs are Hicks-neutral so that we can write $c \equiv \nu/z$ where ν is the marginal cost of a firm with $z = 1$ and z is firm productivity. We show in the Appendix that to guarantee SCD-T it is then sufficient to establish that

$$\varepsilon_{\pi'}(c) \geq 1.$$

Conveniently, this is always satisfies whent the profit functions satisfies SCD-C from below. As a result, whenever $\varepsilon_{\pi'}(c) \geq 1+\theta > 1$, our setup setup satisfies *both* SCD-C from below (complements) and SCD-T. On the other hand, if $1+\theta \geq \varepsilon_{\pi'}(c) \geq 1$, then the firm's profit function satisfies SCD-C from above (substitutes) and SCD-T.

At this juncture, it worth pointing out that the well-known constant elasticity Dixit-Stiglitz demand falls squarely within our framework. In particular, it is easy to show that

$$\varepsilon_{\pi'}(c) = \sigma,$$

where σ is the elasticity of substitution. Therefore, complementarities arise if $\sigma > 1+\theta$ and vice-versa whereas SCD-T holds if $\sigma \geq 1$. Establishing these results allows to use our solution and aggregation method for the benchmark CES case.

4.2 Parameterizing the Demand System

To parameterize the model, we need to put more structure on the demand system. We choose to work with the class of demand systems introduced by Pollak (1971), which have become popular in the literature studying variable markups, e.g., Simonovska (2015), Arkolakis et al. (2019) and Behrens et al. (2020). Our preferred specification is

$$D_n\left(p/P_n^*\right) = \left(p/P_n^*\right)^{-\sigma} + \gamma$$

where $\gamma < 0$.

There are various appealing features of this particular parameterization of demand. First, it features a choke price (Arkolakis et al. (2019)) which implies that entry into each destination market n is guaranteed only for the firms with low enough marginal costs, $c_n \leq P_n^*$. 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 the resulting demand systemimplies that markups increase with firm size a salient finding of recent investigations on the relationship of firm size and firm markups (see De Loecker et al. (2016)).

The price implied by the demand function is given by

$$p\left(c, P_n^*\right) = \frac{\sigma}{(\sigma-1) + \left(p/P_n^*\right)^{\sigma}} c,$$

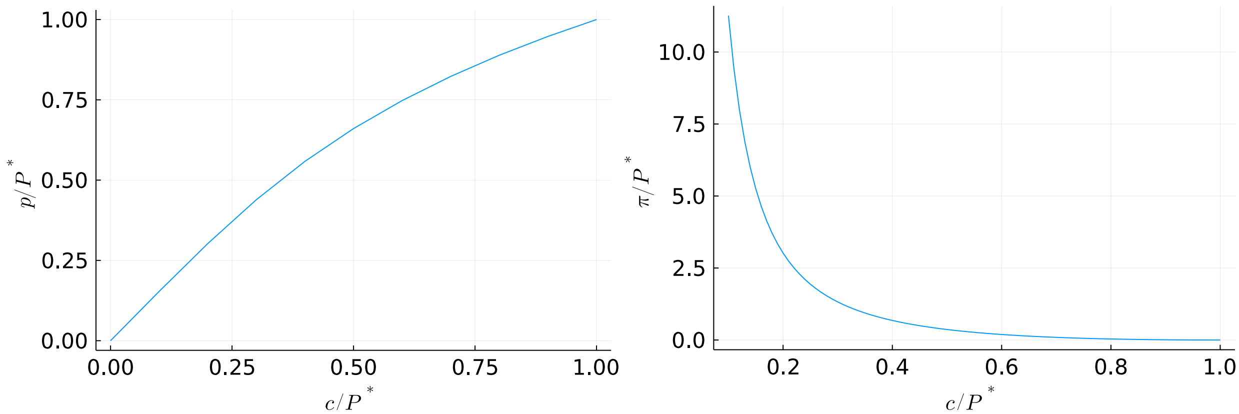
where we notice that the markup is decreasing in the firm marginal cost, and for $p = P^* \iff p = c$, i.e., $c_n^* = P_n^*$ is the maximum production cost that can sustain production in market n. The largest firms in each market charge a constant markup over marginal cost, i.e., as $c \to 0 \implies p \to \frac{\sigma}{\sigma-1} c$. The key parameter σ parsimoniously captures all these effects by regulating the size-elasticity of demand relationship.

Finally, the key elasticity of the derivative of the profit function takes the convenient form

$$\varepsilon_{\pi'} = \sigma \frac{1}{1-(p/P^*)^{\sigma}} \left(\frac{(\sigma-1) + (p/P^*)^{\sigma}}{(\sigma-1) + (\sigma+1)\,(p/P^*)^{\sigma}} \right) \geq \frac{\sigma}{2}.$$

The lower bound $\sigma/2$ makes it straightforward to pick parameters that such that the firm's problem satisfies both SCD-C from below and SCD-T. In particular, we choose σ so that $\frac{\sigma}{2} > 1 + \theta$, always. For the rest of the paper, we assume that σ always takes on a value so that the firm's problem satisfies SCD-C from below.

FIGURE 4: Variable markups with Pollak demand



Notes: Firm outcomes in a single market n with Pollak demand, parameterized by $\sigma \approx 2.86, \gamma = -1$ as in Arkolakis et al. (2019). In contrast to a constant markup setup, firm prices rise more quickly when marginal costs are low than when they are high, translating to higher markups for productive firms compared to unproductive firms. Variable profits are convex in marginal costs.

4.3 Framework Aggregation

Since with the Pollak (1971) the firm's problem satisfies the two single crossing conditions as long as parameters are chosen appropriately, we can apply the solution method to solve numerically for the firm's policy function for each origin i. In this section, we choose a set of reasonable parameters and provide simulation results showcasing our ability to solve the model.

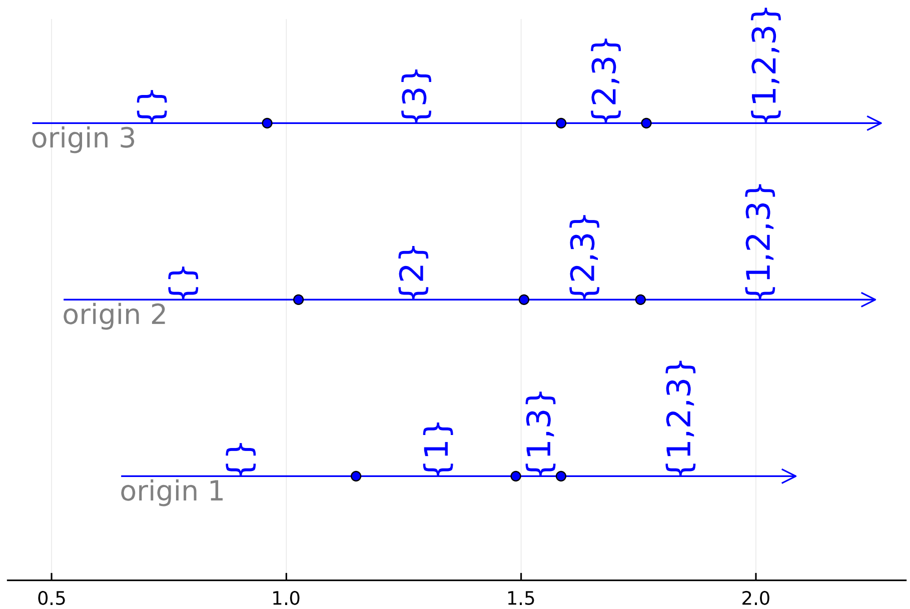
We consider an economy with three countries. First, we set set $\sigma = 2.86$ and $\gamma = -1$ as in Arkolakis et al. (2019). Notice in particular that $\frac{\sigma}{2} > 1 + \theta$ so that both SCD-C and SCD-T are satisfied. Furthermore, we set $\theta = 5$, internal costs to $d_{\ell\ell} = \gamma_{\ell\ell} = 1$, and all external bilateral costs to $\mathbf{d} = 1.56, \boldsymbol{\gamma} = 1.5$. Laslty, fixed cost of plant establishment are symmetrically $f_\ell = 0.1$, wages are symmetrically fixed at $w_\ell = 1$, and total expenditure in each market is fixed at $X_n = n$. Thus, the first market is the smallest, while the third market is the largest.

We compute the equilibrium choke prices P_n^* as the fixed points implied by consumer optimization

$$(P_n^*)^{-\sigma} = \frac{-\gamma \sum_i \int_{\Omega_{in}} p_{in}(\omega)^{1-\sigma} \mathrm{d}\omega}{X_n - \gamma \sum_i \int_{\Omega_{in}} p_{in}(\omega) \mathrm{d}\omega} \qquad \Omega_{in} = \{\omega \mid p_{in}(\omega) \leq P_n^*\}$$

where Ω_{in} is the collection of final goods of firms from country i actually consumed in desti-

Figure 5: Outcomes with endogenous multinational behavior



Notes: Example policy functions for countries symmetric apart for total expenditure, which are set to $X_n = n$. Policy functions for firms from each origin country map firm productivity z on the horizontal axis to optimal location sets. All else equal, market $n = 3$ is the most profitable, since its choke price is the highest. Firms originating there are quick to set up domestic production but face a higher threshold to set up multinational production. On the other hand, firms from the least profitable country $i = 1$ face a high threshold for domestic production, but a comparatively lower one for multinational entry.

nation n. Crucially, final good prices are set given each firm's optimal choice of production locations, so that

$$p_{in}(\omega) = p\left(c_{in}\left(\mathcal{J}_i^*(z); z\right), P_n^*\right),$$

which depends on the the set of plant locations of the firm. Identifying the choke prices therefore entails solving for the policy function multiple times as part of the fixed point procedure.

Given the chosen parameters, we identify the following choke prices, $P_1^* \approx 1.21$, $P_2^* \approx 1.36$, and $P_3^* \approx 1.46$. Since the third market has the highest total expenditure, its choke price is the highest so that even firms with relatively high, i.e., uncompetitive marginal costs produce non-zero quantities. Similarly, the choke price is the lowest in the smallest country which is only served by more competitive firms.

Figure 5 shows the resulting policy functions for firms originating in $i = 1$, $i = 2$, and $i = 3$. Each policy function shows the cutoff productivities at which optimal strategies switch, and the optimal strategies associated with each interval. For example, firms originating in country

1 do not set up any production facilities below productivity ~ 1.15. Above this cutoff, there is an interval of productivity where firms are domestic (that is, they only produce domestically in location 1). The first foreign affiliate is set up in country 3, above productivity ~ 1.49. Finally, the firm expands production to country 2 once productivity exceeds ~ 1.59. A similar interpretation applies for the policy functions of the other two countries of origin. In the presence of arms-length frictions, low-productivity firms produce only domestically while only higher productivity firms are multinationals.

Given this setup, market $n = 3$ is the most profitable from the firm's perspective, since it has the highest choke price. The policy functions reflects this intuition. First, firms in this country face a lower productivity threshold for domestic production. Additionally, there is a large gap between the productivity cutoff for domestic production and multinational production. Opening a foreign production location is valuable since provides better market access to foreign markets, but they are small in our setup. Thus, firms must be very productive before this marginal gain compensates for the fixed cost plant setup. On the other hand, firms originating in countries 1 or 2 do not have the benefit of a relatively profitable home market. Domestic production therefore requires a higher level of productivity. On the other hand, firms originating from both these countries much more readily become multinationals, setting up production in the profitable market $n = 3$ first. By a similar logic, firms originating from the least profitable country, $n = 1$, also set up production in the second most profitable market ($n = 2$) relatively quickly.

4.4 Calibration and Counterfactuals

TBD

5 Conclusion

TBD

References

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*, vol. 4, 55–130, from January 2013.

Arkolakis, C. (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.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2009): "The margins of US trade," *American Economic Review*, 99, 487–93.

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

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

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016): "Prices, markups, and trade reform," *Econometrica*, 84, 445–510.

Eaton, J. and S. Kortum (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

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.

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.

Jiang, E. and V. Tyazhelnikov (2020): "Shadow O shoring and Complementarity in Inputs," .

Keller, W. and S. R. Yeaple (2020): "Multinationals, Markets, and Markups," Tech. rep.

Kimball, M. S. (1995): "The quantitative analytics of the basic neomonetarist model," .

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.

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

Melitz, M. J. and G. I. Ottaviano (2008): "Market size, trade, and productivity," *The review of economic studies*, 75, 295–316.

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.

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.

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.

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

Zheng, F. (2016): "Spatial competition and preemptive entry in the discount retail industry," *Columbia Business School Research Paper*.

6 Single crossing difference conditions

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 monotonicity in set. 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 1 (Sufficiency of sub- and super-nmodularity)**.** *In this first proposition, we show that submodularity and supermodularity are sufficient to ensure monotonicity in set. Fix* $\mathbf{z}$ *and* $\mathbf{y}$ *and consider the return function* π*.*

If π *is submodular, then* π *exhibits monotonicity in set of the substitutes form.*

If π *is supermodular, then* π *exhibits monotonicity in set of the complements form.*

Proof. Since $\mathbf{z}$ and $\mathbf{y}$ are fixed during this proof, they are omitted for notational brevity.

Begin with π submodular. Then, 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 monotonicity in set of the substitutes form must hold. Let $\mathcal{J}_1 \subseteq \mathcal{J}_2$. Select an arbitrary j. The goal is to show that

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

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

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

Then, it is easy to see that applying the submodularity condition implies monotonicity in set of the substitutes form.

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 monotonicity in set of the complements form must hold. Let $\mathcal{J}_1 \subseteq \mathcal{J}_2$. Select an arbitrary j. The goal is to show that

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

Define the sets A and B identically as above for each corresponding scenario. Then, it is easy to see that applying the supermodularity implies monotonicity in set of the complements form.

□

In this second proposition, we show that, conditional on a finite choice space J, monotonicity in set implies sub- or super-modularity.

Proposition 2 (Sufficiency of monotonicity in set with finite choice space)**.** *Fix* $\mathbf{z}$ *and* $\mathbf{y}$ *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 monotonicity in set of the substitutes form, then*

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

If π *exhibits monotonicity in set of the complements form, then*

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

Proof. Since $\mathbf{z}$ and $\mathbf{y}$ are fixed during this proof, they are omitted for notational brevity. Let $\tilde{A}$ and $\tilde{B}$ be arbitrary sets where $\tilde{A} \setminus (\tilde{A} \cap \tilde{B})$ is finite.

Begin with monotonicity in set of the substitutes form first. 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, it is equivalent to show that

$$\pi(I \cup A) + \pi(I \cup B) \geq \pi(I) + \pi(I \cup A \cup B)\,. \tag{5}$$

The proof proceeds inductively on the cardinality of A. When A is empty, then (5) holds with equality.

Now suppose (5) holds for $|A| = n$. Consider the case where $|A| = n + 1$. Let a be an arbitrary element from A and define $\underline{A} \equiv A \setminus \{a\}$ as A with a excluded. 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 monotonicity in set of the substitutes form,

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

Combining the two expressions together yields

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

which confirms (5) for sets A of cardinality $n + 1$. The inductive proof establishes that (5) holds for all A of finite size.

Next, consider monotonicity in set of the complements form. The argument follows a similar structure. Now, it is equivalent to show that

$$\pi(I \cup A) + \pi(I \cup B) \leq \pi(I) + \pi(I \cup A \cup B)\,. \tag{6}$$

Proceed inductively once again on the cardinality of A. When A is empty, (6) holds with equality. Now suppose (6) 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 monotonicity in set of the complements form,

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

Combining the two expressions together yields

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

which confirms (6) for sets A of cardinality $n+1$. The inductive proof establishes that (6) holds for all A of finite size.

□

When $A \setminus (A \cap B)$ is not finite, then monotonicity in set does not necessarily ensure supermodularity or submodularity. 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 j follows as

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

The intuition is simple: since we integrate over the a decision set S for its return, any singular element j is measure zero and has no effect on the decision set's overall return. The return function therefore satsifies monotonicity in set (of both forms).

Now consider $A = [0, 2]$ and $B = [1, 3]$. It is easy to see that

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

so, in this case,

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

Then, when $\alpha > 1$, the return function obeys monotonicity in set of the substitutes form but violates submodularity. Likewise, when $\alpha \in (0, 1)$, the return function obeys monotonicity in set of the complements form but violates supermodularity.

In this next proposition, we establish the sufficient condition for SCD-T provided in the main body of the paper.

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

$$\frac{\partial D_j \pi(\mathcal{J}; \mathbf{z})}{\partial z_i}$$

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

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

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

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

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

and $\mathbf{j}_i$ 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 $Z_j^+(\mathcal{J})$ by construction. In particular, $D_j\pi(\mathcal{J}; \cdot)$ 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 $D_j\pi(\mathcal{J}; \cdot)$ along this dimension is (weakly) positive, the coordinate is increased; otherwise, it is decreased. Thus, $D_j\pi(\mathcal{J}; \cdot)$ 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 $Z_j^+(\mathcal{J})$. Joining these paths together at $\tilde{\mathbf{z}}$, we have constructed a path from $\mathbf{z}$ to $\mathbf{z}'$ that remains in $Z_j^+(\mathcal{J})$. Since $\mathbf{z}$ and $\mathbf{z}'$ were any arbitrary members of $Z_j^+(\mathcal{J})$, we have shown that it is path-connected. Showing $Z_j^-(\mathcal{J})$ is path-connected follows a similar argument. Now suppose $\mathbf{z}$ and $\mathbf{z}'$ are contained in $Z_j^-(\mathcal{J})$. We construct $\tilde{\mathbf{z}}$ in this case as

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

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 $Z_j^-(\mathcal{J})$ by the same logic. $\square$

7 Single Crossing Difference Conditions in Our Multinational Context

In our specific context, signing $\mathrm{d}^2\pi/\mathrm{d}\Theta^2$ is sufficient to establish SCDC. To see that first notice that additional locations are beneficial since they decrease marginal costs. The decision set $\mathcal{J}$ influences marginal costs only through the set's "production potential"

$$\sum_{j\in\mathcal{J}} (\gamma_{ij} w_j d_{jn})^{-\theta} \equiv \sum_{j\in\mathcal{J}} \xi_{jn}$$

(where the i indexing is dropped for now). A location's marginal benefit to a decision set $\mathcal{J}$ can be written as

$$\sum_n \left[f_n\left(\xi_{\ell n} + \sum_{j\in\mathcal{J}} \xi_{jn}\right) - f_n\left(\sum_{j\in\mathcal{J}} \xi_{jn}\right)\right] - F_\ell$$

where $f_n(\Theta) \equiv \pi_n\left(\tilde{\Gamma}\Theta^k/z\right)$ describes variable profits in market n conditional on production potential and F_ℓ is the fixed cost of setting up the location in ℓ. This marginal value is essentially the variable profit gain from increasing production potential by $\xi_{\ell n}$ (when production potential is originally $\sum_{j\in\mathcal{J}} \xi_{\ell n}$), offset by the fixed cost of the location.

The SCDC condition requires that this marginal value crosses zero only once – it is sufficient to show the marginal value is monotonic: that is, given $\mathcal{J}_1 \subseteq \mathcal{J}_2$, the marginal value is bigger (smaller) at $\mathcal{J}_2$ for SCDC from below (above). As a direct consequence of production potential being just a sum across the ξ terms in each decision set, the production potential of $\mathcal{J}_2$ exceeds the production potential of $\mathcal{J}_1$:

$$\sum_{j\in\mathcal{J}_1} \xi_{jn} \leq \sum_{j\in\mathcal{J}_2} \xi_{jn}$$

so it suffices to show that each f_n has increasing (decreasing) differences for SCDC from below (above). In other words, we are comparing the variable profit gain from increasing production potential by $\xi_{\ell n}$ (adding location ℓ) when starting at the larger original production potential (associated with the larger set $\mathcal{J}_2$) to starting at the smaller original production potential (associated with the smaller set $\mathcal{J}_1$).

Formally, we can write the marginal value as

$$\sum_n \int_0^{\xi_{\ell n}} \left[\frac{\partial}{\partial y} f_n\left(y + \sum_{\mathcal{J}} \xi_{jn}\right)\right] \mathrm{d}y - F_\ell = \sum_n \int_0^{\xi_{\ell n}} f_n'\left(y + \sum_{\mathcal{J}} \xi_{jn}\right) \mathrm{d}y - F_\ell$$

as long as the derivative exists. Comparing this marginal value across two decision sets,

$$\sum_n \int_0^{\xi_{\ell n}} \left[f_n'\left(y + \sum_{\mathcal{J}_2} \xi_{jn}\right) - f_n'\left(y + \sum_{\mathcal{J}_1} \xi_{jn}\right)\right] \mathrm{d}y$$

$$= \sum_n \int_{\sum_{\mathcal{J}_1}\xi_{jn}}^{\sum_{\mathcal{J}_2}\xi_{jn}} \frac{\partial}{\partial x}\left[\int_0^{\xi_{\ell n}} f_n'(x+y)\mathrm{d}y\right] \mathrm{d}x = \sum_n \int_{\sum_{\mathcal{J}_1}\xi_{jn}}^{\sum_{\mathcal{J}_2}\xi_{jn}} \left[\int_0^{\xi_{\ell n}} f_n''(x+y)\mathrm{d}y\right] \mathrm{d}x$$

provided the second derivative exists. Thus, $f_n'' \geq 0$ is sufficient to guarantee the sign of the RHS is positive, i.e. the marginal value of ℓ to the larger set $\mathcal{J}_2$ exceeds its marginal value ot the smaller set $\mathcal{J}_1$. Then, we are guaranteed SCDC from below. Similarly, $f_n'' \leq 0$ is sufficient to guarantee SCDC from above.

The CES-Frechet setup of Antras et al. (2017) is a simple example of this argument. The profit function in that framework is

$$\pi(c(\mathcal{J})) = \sum_n \frac{X_n}{\sigma}\left(\frac{\sigma}{\sigma-1}\frac{\tilde{\Gamma}}{zP_n}\right)^{1-\sigma}\left(\sum_{j\in\mathcal{J}} \xi_{\ell n}\right)^{\frac{\sigma-1}{\theta}} - \sum_{\mathcal{J}} F_j$$

so

$$f_n(x) = \frac{X_n}{\sigma}\left(\frac{\sigma}{\sigma-1}\frac{\tilde{\Gamma}}{zP_n}\right)^{1-\sigma} x^{\frac{\sigma-1}{\theta}}$$

whose second derivative is positive when $(\sigma-1)/\theta > 1$ (i.e. the function is convex) and negative when $(\sigma-1)/\theta < 1$ (i.e. the function is concave). These conditions are the exact conditions derived in the original Antras et al. (2017) paper that guarantee supermodularity and submodularity, respectively.

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

$$c_n(\mathcal{J};z)=\frac{\nu(\mathcal{J})}{z}=\frac{\tilde{\Gamma}}{z}\left[\sum_{j\in\mathcal{J}}\xi_{jn}\right]^{-\frac{1}{\theta}}$$

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

$$\pi\left(\frac{\nu(\mathcal{J}\cup\ell)}{z}\right)-\pi\left(\frac{\nu(\mathcal{J})}{z}\right)-F_\ell=\int_{\frac{\nu(\mathcal{J})}{z}}^{\frac{\nu(\mathcal{J}\cup\ell)}{z}}\pi'(c)\mathrm{d}c-F_\ell=\int_{\nu(\mathcal{J})}^{\nu(\mathcal{J}\cup\ell)}\pi'\left(\frac{\nu}{z}\right)\frac{1}{z}\mathrm{d}\nu-F_\ell$$

via a change in variables. Since $\nu$is lower at $\mathcal{J}\cup\ell$, we finally have

$$\pi\left(\frac{\nu(\mathcal{J}\cup\ell)}{z}\right)-\pi\left(\frac{\nu(\mathcal{J})}{z}\right)-F_\ell=\int_{\nu(\mathcal{J}\cup\ell)}^{\nu(\mathcal{J})}\left[-\pi'\left(\frac{\nu}{z}\right)\right]\frac{1}{z}\mathrm{d}\nu-F_\ell$$

where note of course $\pi'\leq 0$ (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 it's monotonic in productivity. Comparing the marginal value at two values of productivity $z_1\leq z_2$,

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

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

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

so it is sufficient to show this difference is positive. Note that the first square bracketed term is positive. Then, it is sufficient for $\varepsilon_{\pi'}(c)\geq 1$ for SCDT to hold.

8 The 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 j such that $j\in\overline{\mathcal{J}}^{(K)}\setminus\underline{\mathcal{J}}^{(K)}$ and forms two subproblems, or "branches:" one in which j is included in $\mathcal{J}^\star$ (i.e., added to $\underline{\mathcal{J}}$) and one in which it is excluded from $\mathcal{J}^\star$ (i.e., excluded from $\overline{\mathcal{J}}$).[27] The two fixed points resulting from applying the squeezing procedure to the bounding

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

sets of each subproblem are the optimal decision sets *conditional* on the assumed inclusion or exclusion of j. 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 j' and create two sub-sub-problems. Recursively applying the branching step in such a way creates a "tree," where the terminal nodes are subproblems for which the squeezing procedure has converged to a bounding set pair where the lower bound and upper bound are equal.

We now formally define the branching step.

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

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

The mapping B^b is given by

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

For given initial bounding sets $[\underline{\mathcal{J}}, \overline{\mathcal{J}}]$, we denote the operator of applying the branching step until global convergence by $B^{a(K)}([\underline{\mathcal{J}}, \overline{\mathcal{J}}])$ and $B^{b(K)}([\underline{\mathcal{J}}, \overline{\mathcal{J}}])$, respectively. Global convergence of the branching step occurs when the stopping condition $\underline{\mathcal{J}} = \overline{\mathcal{J}}$ is met on each branch.[28]

Suppose the return function exhibits SCD-C from above.[29] Given an initial bounding pair $[\underline{\mathcal{J}}, \overline{\mathcal{J}}]$ with $\underline{\mathcal{J}} \subseteq \mathcal{J}^\star \subseteq \overline{\mathcal{J}}$, the globally converged result $B^{a(K)}([\underline{\mathcal{J}}, \overline{\mathcal{J}}])$ 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 **??** which shows an example of a tree created by the branching procedure. It starts with a bounding set

[28]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.

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

pair for which the squeezing procedure has converged, but there still remain undetermined items. One of these, j, is selected. Two branches based on this item are formed. The left hand branch corresponds to the subproblem where j is presumed to be excluded from the optimal decision set, while the right hand branch corresponds to the subproblem where j 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 $\underline{\mathcal{J}} = \overline{\mathcal{J}}$, yielding the orange $\mathcal{J}$. This decision set is optimal precisely conditional on the requirement that j must be included. On the other hand, convergence of the squeezing procedure in the left hand branch does not deliver an optimal decision set. The returned bounding pair is still such that there are strictly more items in the upper bound than lower bound set. The branching procedure therefore branches again, this time selecting the still undetermined item j' 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 $\mathcal{J}$. This green decision set $\mathcal{J}$ is optimal conditional on both j and j' 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 j''. This branch yields conditionally optimal decision sets, the brown and pink $\mathcal{J}$s. The first is optimal conditional on j, j', and j'' all being excluded. Likewise, the second is optimal conditional on excluding j and j', but including j''. 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 2 and let the bounding pair $[\underline{\mathcal{J}}, \overline{\mathcal{J}}]$ be such that $\underline{\mathcal{J}} \subseteq \mathcal{J}^\star \subseteq \overline{\mathcal{J}}$.

Then, if π exhibits SCD-C from above,

$$\mathcal{J}^* = \underset{\mathcal{J} \in B^{a(K)}(S^a([\underline{\mathcal{J}}, \overline{\mathcal{J}}]))}{\arg\max} \pi(\mathcal{J}) \, .$$

while if π exhibits SCD-C from below,

$$\mathcal{J}^* = \underset{\mathcal{J} \in B^{b(K)}(S^b([\underline{\mathcal{J}}, \overline{\mathcal{J}}]))}{\arg\max} \pi(\mathcal{J}) \, .$$

9 Generalized Branching Procedure

We define the generalized branching step as follows.

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

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

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

The mapping B^b is given by

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

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

Given an initial 4-ple $[(\underline{\mathcal{J}}, \overline{\mathcal{J}}, M), Z]$ and an undetermined item $j \in M$, the generalized branching step creates two branches or subproblems. The first supposes that j 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-ples which exhaustively partition the original region Z.[30] Each branch may now contain different partitions of the original typespace.

On either branch, if there are any 4-ples 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-ple when the bounding sets for that 4-ple coincide. Global convergence occurs when bounding sets coincide for 4-ples on all branches. Then, the output of the generalized branching procedure is a collection of 4-ples each of the form $[\mathcal{J}, \mathcal{J}, \emptyset, Z]$.

For illustration, Figure **??** depicts the process of applying the generalized branching procedure to an initial 4-ple $[(\underline{\mathcal{J}}, \overline{\mathcal{J}}, M), Z]$. The initial 4-ple specifies lower and upper bound sets $(\underline{\mathcal{J}}, \overline{\mathcal{J}})$ 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 j to be excluded, and the other requiring that item j 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 j. On the left branch, j 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, j'. This procedure repeats until

[30] 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.

no undetermined elements remain in any of the branches for any type $z \in Z$.

Now consider the entire initial region Z, repeated at the bottom of the graphic. The repeated application of the squeezing procedure to smaller and smaller subregions of the typespace creates subregions of the overall typespace 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 $[\underline{z}, \overline{z}]$.

10 Solution results

In this section, we discuss results related to our solution methods and the policy function. We begin by proving Theorems 1 and 2 before closing with characterisations of the policy function under SCD-C.

Proof. We start with Theorem 1. Suppose the return function satisfies SCD-C from above. We prove inductively that successively applying the squeezing step weakly narrows down the choice space without eliminating the optimal decision set. Let the bounding sets after the nth application be $[\underline{\mathcal{J}}^{(k)}, \overline{\mathcal{J}}^{(k)}]$.

Starting with $[\emptyset, J]$, it is trivially the case that $\emptyset \subseteq \underline{\mathcal{J}}^{(1)}$ and $\overline{\mathcal{J}}^{(1)} \subseteq J$. What remains to show is that this first pair of bounding sets sandwiches $J^\star$. Let $j \in \underline{\mathcal{J}}^{(1)}$. Then, $D_j\pi(J) > 0$ so $D_j\pi(\mathcal{J}^\star) > 0$ by SCD-C from above. So $j \in \mathcal{J}^\star$. Since j was an arbitrary element of $\underline{\mathcal{J}}^{(1)}$, we conclude $\underline{\mathcal{J}}^{(1)} \subseteq \mathcal{J}^\star$. Next, consider $j \in \mathcal{J}^\star$. Then, $D_j\pi(\mathcal{J}^\star) > 0$ by optimality, so $D_j\pi(\emptyset) > 0$ by SCD-C from above. Thus, $j \in \overline{\mathcal{J}}^{(1)}$ and since j was an arbitrary member of $\mathcal{J}^\star$, it must be the case that $\mathcal{J}^\star \subseteq \underline{\mathcal{J}}^{(1)}$.

Now suppose $\underline{\mathcal{J}}^{(n-1)} \subseteq \underline{\mathcal{J}}^{(n)} \subseteq \mathcal{J}^\star \subseteq \overline{\mathcal{J}}^{(n)} \subseteq \overline{\mathcal{J}}^{(n-1)}$. We show that $\underline{\mathcal{J}}^{(n)} \subseteq \underline{\mathcal{J}}^{(n+1)} \subseteq \mathcal{J}^\star \subseteq \overline{\mathcal{J}}^{(n+1)} \subseteq \overline{\mathcal{J}}^{(n)}$. Select $j \in \underline{\mathcal{J}}^{(n)}$. It must be the case that $D_j\pi(\overline{\mathcal{J}}^{(n-1)}) > 0$. Since $\overline{\mathcal{J}}^{(n)} \subseteq \overline{\mathcal{J}}^{(n-1)}$, SCD-C from above implies $D_j\pi(\overline{\mathcal{J}}^{(n)}) > 0$, so $j \in \underline{\mathcal{J}}^{(n+1)}$. Since j was an arbitrary element of $\underline{\mathcal{J}}^{(n)}$, we conclude $\underline{\mathcal{J}}^{(n)} \subseteq \underline{\mathcal{J}}^{(n+1)}$. Similarly, now select $j \in \mathcal{J}^{(n+1)}$. We show it is in $\mathcal{J}^{(n)}$. Because $D_j(\underline{\mathcal{J}}^{(n)}) > 0$ and $\underline{\mathcal{J}}^{(n-1)} \subseteq \underline{\mathcal{J}}^{(n)}$, SCD-C from above ensures that that $D_j(\underline{\mathcal{J}}^{(n-1)}) > 0$. We conclude $\mathcal{J}^{(n+1)} \subseteq \mathcal{J}^{(n)}$.

We now show $\underline{\mathcal{J}}^{(n+1)}$ and $\overline{\mathcal{J}}^{(n+1)}$ sandwich the optimal decision set. Let $j \in \underline{\mathcal{J}}^{(n+1)}$ so that $D_j(\overline{\mathcal{J}}^{(n)}) > 0$. By the inductive assumption, $\mathcal{J}^\star \subseteq \mathcal{J}^{(n)}$, so SCD-C from above allows us to conclude that $D_j(\mathcal{J}^\star) > 0$, implying $j \in \mathcal{J}^\star$ optimally. Similarly, suppose $j \in \mathcal{J}^\star$ so that $D_j(\mathcal{J}^\star) > 0$. By the inductive assumption, $\underline{\mathcal{J}}^{(n)} \subseteq \mathcal{J}^\star$, so SCD-C from above implies $D_j(\underline{\mathcal{J}}^{(n)}) > 0$, ensuring $j \in \underline{\mathcal{J}}^{(n+1)}$.

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

Finally, the squeezing procedure must complete in under $|J|$ 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 $|J|$ items eventually being fixed. $\square$

Having proven Theorem 1, the proof for Theorem 2 follows similarly.

Proof. Consider the 4-ple $[(\underline{\mathcal{J}}, \overline{\mathcal{J}}, M), Z]$ 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-ples induce functions $\underline{\mathcal{J}}(\cdot)$ and $\overline{\mathcal{J}}(\cdot)$ over Z. We then show $\underline{\mathcal{J}} \subseteq \underline{\mathcal{J}}(\mathbf{z}) \subseteq \mathcal{J}^\star(\mathbf{z}) \subseteq \overline{\mathcal{J}}(\mathbf{z}) \subseteq \overline{\mathcal{J}}$ for every $\mathbf{z} \in Z$.

First, observe that $Z_j^+(\overline{\mathcal{J}})$ and $Z_j^-(\underline{\mathcal{J}})$ are disjoint. For any type $\mathbf{z}$ receiving positive benefit from j's addition to $\overline{\mathcal{J}}$, SCD-C from above implies that it must receive positive benefit from j's addition to $\underline{\mathcal{J}}$. Then, $Z_j^0(\underline{\mathcal{J}}) \cup Z_j^+(\underline{\mathcal{J}})$ is the complement of $Z_j^-(\underline{\mathcal{J}})$, from which $Z_j^-(\underline{\mathcal{J}})$ has been removed for the third 4-ple. The three new 4-ples thus induce a valid partitioning on Z, inducing the functions $\underline{\mathcal{J}}(\cdot)$ and $\overline{\mathcal{J}}(\cdot)$.

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

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

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

Proposition 4 (Nested policy function on single dimensional typespace)**.** *Consider the satisfying SCD-C from below and SCD-T where the choice set J is finite. Then, for any $z_1 < z_2$, it must be that $\mathcal{J}^\star(z_1) \subseteq \mathcal{J}^\star(z_2)$.*

Proof. For a contradiction, suppose not. Define $\mathcal{J}^o \equiv \mathcal{J}^\star(z_1) \setminus \mathcal{J}^\star(z_2)$, which has cardinality $N \in (0, \infty)$ by assumption. Further, let $\mathcal{J}^i \equiv \mathcal{J}^\star(z_1) \cap \mathcal{J}^\star(z_1)$ so that $\mathcal{J}^\star(z_1) = \mathcal{J}^i \cup \mathcal{J}^o$.

Claim: for any finite N, it must be that $\mathcal{J}^\star(z_2) \cup \mathcal{J}^o$ is preferable to $\mathcal{J}^\star(z_2)$ for z_2 types. Note $\mathcal{J}^o \neq \emptyset$, so $\mathcal{J}^\star(z_2) \cup \mathcal{J}^o \neq \mathcal{J}^\star(z_2)$. This claim represents a contradiction to $\mathcal{J}^\star(z_2)$ being the decision set for z_2. The proof for the claim proceeds by induction on N. Suppose $N = 1$. Let its element be j. Then,

$$\begin{gathered} 0 < D_j \pi(\mathcal{J}^\star(z_1), z_1) = D_j \pi(\mathcal{J}^i, z_1) \\ \Longrightarrow 0 < D_j \pi(\mathcal{J}^i, z_2) \\ \Longrightarrow 0 < D_j \pi(\mathcal{J}^\star(z_2), z_2) \\ \Longrightarrow \pi(\mathcal{J}^\star(z_2) \cup \{j\}, z_2) > \pi(\mathcal{J}^\star(z_2), z_2) \end{gathered}$$

where the first line derives from z_1 optimality[31], the second from single-dimensional SCDT, and the third SCD-C from above, and the fourth from recognising that $j \notin \mathcal{J}^\star(z_2)$. This argument proves the claim for $N = 1$.

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

$$\pi(\mathcal{J}^\star(z_2), z_2) < \pi(\mathcal{J}^\star(z_2) \cup \mathcal{J}_n^o, z_2)\,.$$

Using the optimality of z_1 types, SCDT, and SCDC-B,

$$\begin{gathered} 0 < D_j \pi(\mathcal{J}^\star(z_1), z_1) = D_j \pi(\mathcal{J}^i \cup \mathcal{J}_n^o; z_1) \\ \Longrightarrow 0 < D_j \pi(\mathcal{J}^i \cup \mathcal{J}_n^o; z_2) \\ \Longrightarrow 0 < D_j (\mathcal{J}^\star(z_2) \cup J_n^o; z_2) \\ \Longrightarrow \pi(\mathcal{J}^\star(z_2) \cup \mathcal{J}_n^o \cup \{j\}, z_2) > \pi(\mathcal{J}^\star(z_2) \cup \mathcal{J}_n^o, z_2) \geq \pi(\mathcal{J}^\star(z_2), z_2) \end{gathered}$$

[31] To break ties, we assume that all items for which an agent is indifferent are excluded from the the optimal set. These are easily identified and can be included into optimal sets if desired.

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 $\mathcal{J}^o$, agents of type z_2 prefer $\mathcal{J}^\star(z_2) \cup \mathcal{J}^o$ to $\mathcal{J}^\star(z_2)$, a contradiction. $\square$

We extend the nesting result to the multidimensional setting in this next proposition, with a stricter condition in place of SCD-T.

Proposition 5 (Nested policy function on multidimensional typespace)**.** *Suppose that the return function satisfies SCD-C from below and the choice set J is finite. In addition, suppose that, for each component i of the type vector,*

$$\frac{\partial D_j \pi(\mathcal{J}; \mathbf{z})}{\partial z_i}$$

maintains the same sign for all j, $\mathcal{J}$, 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, $\mathcal{J}^\star(\mathbf{z}_1) \subseteq \mathcal{J}^\star(\mathbf{z}_2)$.

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

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

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 j's addition in $\mathcal{J}$ than type $\mathbf{z}_1$.

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

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

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

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

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

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 $\mathcal{J}^\star(\mathbf{z}_2) \cup \mathcal{J}^o_n \cup \{j\}$ to $\mathcal{J}^\star(\mathbf{z}_2) \cup \mathcal{J}^o_n$, which is itself preferred to $\mathcal{J}^\star(\mathbf{z}_2)$ by the inductive assumption.

We have shown that, for any finite non-empty $\mathcal{J}^o$, agents of type $\mathbf{z}_2$ prefer $\mathcal{J}^\star(\mathbf{z}_2) \cup \mathcal{J}^o$ to $\mathcal{J}^\star(\mathbf{z}_2)$, a contradiction. □

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 for each i, j, and $\mathcal{J}$, as long as it is the same across the typespace given these. On the other hand, the condition provided for the multidimensional nesting result requires the sign to remain the same for each i, regardless of which j and $\mathcal{J}$ are chosen.

11 Input sourcing and value chain framework

We first show that the unit cost of production in our GVC framework is proportional to $\Theta_m(\{\mathcal{J}^k\}_k)^{-1/\theta}$. We then show that an alternative framework where each intermediate is a "snake" yields the same firm decisions and allocations as our baseline framework.

Begin by considering each destination market separately since the firm's activities in one do not affect the firm's outcomes in another. Fix a destination market m and proceed by induction on the number of stages K. Begin first with $K = 2$ and consider the unit cost implied by a given sourcing strategy $\{j^1(\omega^1), j^2(\omega^2)\}$. The first stage is produced with labor only, so the cost of producing any intermediate ω^1 is $w_{j^1(\omega^1)}$. Then, the total unit cost conditional on the sourcing strategy is

$$\exp\left\{\int\int \ln\left[\frac{d_{j^2(\omega^2)m}}{\nu(\omega^2)}\left(w_{j^1(\omega^1)}d_{j^1(\omega^1)j^2(\omega^2)}\right)^{\alpha^2} w_{j^2(\omega^2)}^{1-\alpha^2}\right]\mathrm{d}\omega^1\mathrm{d}\omega^2\right\}$$

where we have used the Cobb-Douglas structure of production.

For the inductive step, we assert that unit costs given a sourcing strategy are

$$\exp\left\{\int \ln\left[\frac{1}{\nu(\omega^K)}\prod_{k=1}^{K} w_{j^k(\omega^k)}^{\beta^k} d_{j^k(\omega^k)j^{k+1}(\omega^{k+1})}^{\frac{\alpha^{k+1}}{1-\alpha^{k+1}}\beta^{k+1}}\right]\mathrm{d}\omega^1\dots\mathrm{d}\omega^K\right\}$$

for a production structure with K stages. We now show the statement holds for production structures with $K+1$ stages. Let this expression omitting the shipping shock be denoted $\tilde{a}_m$ and consider the final stage of production, where a unit continuum of ω^K intermediates must be sourced for each ω^{K+1}. The cost of doing so is $\tilde{a}_{j^{K+1}(\omega^{K+1})}$. Using the properties of Cobb-Douglas production, the final unit cost given a sourcing strategy will therefore be

$$\begin{aligned}
&\exp\left\{\int \ln\left[\frac{d_{j^{K+1}(\omega^{K+1})m}}{\nu(\omega^{K+1})} w_{j^{K+1}(\omega^{K+1})}^{1-\alpha^{K+1}}\left(\tilde{a}_{j^{K+1}(\omega^{K+1})}\right)^{\alpha^{K+1}}\right]\mathrm{d}\omega^{K+1}\right\}\\
=\ &\exp\left\{\int \ln\left[\frac{1}{\nu(\omega^K+1)}\prod_{k=1}^{K+1} w_{j^k(\omega^k)}^{\beta^k} d_{j^k(\omega^k)j^{k+1}(\omega^{k+1})}^{\frac{\alpha^{k+1}}{1-\alpha^{k+1}}\beta^{k+1}}\right]\mathrm{d}\omega^1\dots\mathrm{d}\omega^K\right\}
\end{aligned}$$

completing the inductive step.

What remains is to characterize the unit costs of the firm once it has selected the optimal sourcing strategy. Since there is a continuum of final-stage intermediates ω^K, the Fréchet distribution of shipping shocks $\nu(\omega^K)$ is exactly realized at the firm level. Thus, for each intermediate ω^K, a shipping strategy is chosen conditional on the shock draws

$$\begin{aligned}
a_m(\{\mathcal{J}^k\}_k) &= \exp\left\{\int \ln\left[\min_{\{j^k(\omega^k)\}_k}\ \nu(\omega^K)\prod_{k=1}^{K} w_{j_m^k(\omega^k)}^{\beta^k} d_{j_m^k(\omega^k)j_m^{k+1}(\omega^{k+1})}^{\frac{\alpha^{k+1}}{1-\alpha^{k+1}}\beta^{k+1}}\right]\mathrm{d}\omega^1\dots\mathrm{d}\omega^K\right\}\\
a_m(\{\mathcal{J}^k\}_k) &\equiv \exp\left\{\mathbb{E}\left[\ln M\right]\right\}\\
&= \exp\left\{\ln\Theta_m(\{\mathcal{J}^k\}_k)^{-\frac{1}{\theta}} - \frac{\gamma}{\theta}\right\} \propto \Theta_m(\{\mathcal{J}^k\}_k)^{-\frac{1}{\theta}}
\end{aligned}$$

where M denotes the minimum. This minimum is distributed Weilbull with shape θ and scale $\Theta_m(\{\mathcal{J}^k\}_k)^{-1/\theta}$, given the shocks are Fréchet distributed. The third line follows from the properties of the Weilbull distribution.

We now explore an alternative framework where each ω^K intermediate is instead produced with a snake-structure, so that each ω^K corresponds to the path of an initial intermediate ω^1 being progressively transformed as it moves through the value chain. In particular, suppose that the previous stage input is combined Cobb-Douglas with labor each stage. The first stage remains produced with labor only. The firm's total production is then

$$q_m = \exp\left\{\int \ln\left[\frac{\nu^K(\omega^K)}{d_{j^K(\omega^K)m}} q^K(\omega^K)\right] \mathrm{d}\omega^K\right\} \quad q^k(\omega^K) = \left(\frac{q^{k-1}(\omega^K)}{d_{j^{k-1}(\omega^K)j^k(\omega^K)}\alpha^k}\right)^{\alpha^k} \left(\frac{\ell^k(\omega^k)}{1-\alpha^k}\right)^{1-\alpha^k}$$

given a path $\{j^k(\omega^K)\}_j$ for each intermediate ω^K. Through a similar inductive argument to above, the unit cost of production given a set of paths is

$$\exp\left\{\int \ln\left[\frac{1}{\nu(\omega^K)} \prod_{k=1}^{K} w_{j^k(\omega^K)}^{\beta^k} d_{j^k(\omega^K)j^{k+1}(\omega^K)}^{\frac{\alpha^{k+1}}{1-\alpha^{k+1}}\beta^{k+1}}\right] \mathrm{d}\omega^K\right\}$$

where the β^ks once again represent the overall weight on labor from each stage. We now assume the shipping shocks $\nu(\omega^K)$ for an intermediate ω^K are drawn for each potential path, with the firm choosing the path yielding the lowest cost. Once again appealing to the law of large numbers, the Fréchet distribution of shocks is realized across the continuum of intermediates ω^K. In a similar argument to above, the unit cost of production is therefore identical to the one derived above. Moreover, on aggregate, each location-stage $j \in \mathcal{J}^k$ carries out the same amount of production in both frameworks. This share is encapsulated described by their contribution to the network potential term, as is common in gravity models with a multilateral access term.