Location Choices of Multi-plant Oligopolists

Chenying Yang*

Singapore Management University

August 2021
(Click here for the latest version)

Abstract

I develop a quantitative model of multi-plant oligopolists where each firm decides where to locate the set of plants and how to serve each market, taking into account cannibalization across its own plants as well as competition with others. In contrast to canonical trade models with multinational firms where neither spatial interdependency of decisions nor oligopoly is considered, I advance the existing research by allowing for interdependent entry, oligopolistic rivalry and variable markups. Despite having a high-dimensional discrete choice problem, I provide an estimation toolkit for the model in a three-step procedure, leveraging the gravity-type regressions, the analytical expression for market price given the spatial distribution of plants, and the solution algorithm for a combinatorial problem when the location game is submodular. I present simulation-based evidence to show that neglecting interdependencies among plant locations within a multi-plant firm introduces quantitatively relevant differences in estimation.

Keywords: Multi-plant, oligopoly, interdependent entry, combinatorial discrete choice, submodular games.

*Please send any comments to cyyang@smu.edu.sg

1 Introduction

Evaluating how firms, especially multi-plant firms, operate spatially is a complex problem. A multi-plant firm confronts trade-offs in deciding where to locate a set of plants and which plant to supply where, taking into account the competition with rival firms. A firm having too many plants close to consumers incurs higher fixed costs. Having too sparse plants implies higher costs for transporting products to consumers. Further, given locations differ in production costs and plants can be heterogeneous in productivity, plants owned by the same firm would cannibalize each other's markets. The possibility of trade and cannibalization causes the plant entry decision in one location interdependent of the other, which the problem becomes combinatorial. The geographic distribution of plants subsequently determines the goods flow and prices in each market.

In this paper, I develop a quantitative spatial model of heterogeneous oligopolists to study their decisions associated with multi-plant production. It is firmly grounded at the intersection of trade and industrial organization theory. With this model, I aim to answer three main questions. First, what determines the number and location of plants for different firms? Second, how do markups and prices vary by the spatial allocation of plants? Third, how important is allowing for multi-plant production and interdependent entry of plants?

To shed light on these essential components of firms' decisions, I set up an economy with a finite number of discrete locations that are heterogeneous in productivities and input costs. Plants at each location are potential suppliers of local consumers and those in every other locations. The productivity of a plant depends on local characteristics as well as a stochastic term that is realized post entry. Firms decide how many and which locations to build plants by balancing their expectation on costs of production, costs of distribution and costs of entry, while factoring competition within and across firms given the vector of demand. When competing, plants engage in head-to-head price competition except that those owned by the same firm coordinate. Therefore, for a multi-plant firm, there are counteracting forces in determining the optimal set of production locations: coordination in prices set by the firm improves its competitive advantage against rivals through building more and favorably located plants, whereas cannibalization between its own plants decreases the marginal benefit. Plants are strategically added until marginal payoff cannot cover fixed costs of building. Across firms, plants are strategic substitutes.

The model advances existing research in two aspects. First, firms are granular and oligopolies such that the model yields firm-market-specific variable markups. Existing work has long been using constant markup derived from the workhorse model of constant elasticity of substitution (CES) with monopolistic competition.[1] I relax those assumptions and incorporate strategic pricing among

[1]Examples of using Dixit-Stiglitz type of model in studying multinational firm are Tintelnot (2017), Antras et al. (2017), and Hu and Shi (2019).

oligopolistic rivals, while keeping the markup distribution tractable with an analytical expression. The key idea for tractability in the multi-plant firm framework is to consider a nested structure in competition where plants within a firm do not undercut each other in price but plants across firms compete head-to-head similar to Bernard et al. (2003) (hereforce BEJK). By allowing the particular competition structure, a standard parametric distributional assumption of plant productivities, and a continuum of consumers in each market, I align the multi-plant firm model in this paper to many other canonical models in the literature that conform to a gravity trade framework. It delivers a closed-form expression for firms' expected sales given the plant location configuration.

The model predicts that a firm with more and better located plants charges higher markups and captures a larger fraction of the market. The intuition is that favorable plant bundle a firm owned lowers its average costs and improves the firm's competitive advantage to outbid others. Introduction of entry and oligopolistic competition opens up the possibility to reexamine the connection between extensive and intensive margin in the context of multi-plant or broadly multinational firms. It highlights the role of spatial distribution of plants in trade besides the usual technological differences and geographic barriers. As the geographic configuration of plants shapes the relative competency of firms, it also reinforces differences in comparative advantage across locations.

Second, the plant location game is interdependent and submodular, and yet a model with such feature can be fully estimated. Solving the interdependent entry game facing a large set of potential locations implies a hard permutation problem. When there are L number of possible production locations, a firm faces 2^L possible choices. A game with F number of players further complicates the combinatorial discrete choice (CDC) problem as it now involves 2^{FL} combinations. Due to the difficulty in tackling such a problem, some papers either assume away export platform sales or fixed costs of entry, or treat locations as predetermined.[2] There are additional theoretical challenges with respect to the existence and uniqueness of high dimensional spatial equilibrium when players are strategic substitutes.[3] In games with strategic complements, it follows from Topkis' theorem (Topkis, 1978) and Tarski's fixed point theorem (Tarski et al., 1955) that a pure strategy Nash equilibrium (PSNE) exists. For this reason, most papers that deal with combinatorial problems are applications of games with strategic complements, such as Jia (2008) where chain stores exhibit positive spillover and Alfaro et al. (2021) where firms prefer to export to similar markets. However, little is known about games with strategic substitutes. One could accommodate by not solving the game but using a revealed preference approach to bound the model primitives (Holmes, 2011).

[2]Examples are Helpman et al. (2004), Ramondo and Rodríguez-Clare (2013), Irarrazabal et al. (2013), and Head and Mayer (2019).

[3]I take definitions from Jackson and Zenou (2015) p.103, where a game has *strategic complements* when "an increase in the actions of other players leads a given player's higher actions to have relatively higher payoffs compared to that player's lower actions". On the opposite, games of *strategic substitutes* are "an increase in other players' actions leads to relatively lower payoffs to higher actions of a given player".

In this paper, I leverage the submodularity and aggregation property of the firm's profit function to solve the PSNE of a game with imperfect competition by applying a CDC solution algorithm designed in Eckert et al. (2017). I then provide a toolkit to separately identify model primitives in three steps using rather aggregated and easily obtained data. Conditions on key parameters to guarantee the existence of a PSNE can be validated in steps prior to solving the location game.

Specifically, the estimation procedure is the following. In the first step, gravity regressions and data on bilateral trade are used to estimate a composite of local TFP and input costs that determines the relative production advantage across locations. I can also obtain an estimate of trade elasticity which regulates competition intensity among plants. In the second step, I estimate demand through nonlinear least squares using data on consumption and market characteristics. Instead of using price data which is subject to limited availability, I use the model-derived price index constructed from the parameters recovered in the last step. It is a function of the every potential suppliers' costs and markup rule, weighted by its probability of serving the market. In the third step, I estimate the fixed costs of building plants by fitting moments using the observed plant locations. A notable advantage is that the multi-plant firm model in this paper can be estimated with minimum data requirement. Micro data on firm or plant-level market shares or prices is not needed for this exercise.

Although I present in this paper a model characterizing a rich set of multi-plant firms' decisions and also provide a recipe to recover it without sacrificing too much on computational cost when one has limited data, a researcher may still worry about the payoff for incorporating interdependent entry when studying multi-plant firms. After all, one could assume every plant entry decision is isolated, and thus, a firm decides to have a plant if and only if the expected variable profit associated from the particular plant is higher than its fixed cost of establishing. The assumption is not as unreasonable as it appears to be. Think of every geographical unit is managed by a manager who has complete authority to decide whether to have a plant and how much to charge, and the headquarter has no influence on any decisions. In that case, plants and firms are conceptually equivalent. Here, I essentially take the other extreme: each firm has a big managing board who oversees entry and pricing. The reality could be somewhere in between. Instead of arguing which premise is correct, the focus is whether we can approximate multi-plant firm decisions using single-plant assumptions if the true data generating process (DGP) is the CDC model with oligopolistic multi-plant pricing. I present simulation evidences that ignoring interdependent entry generates remarkable biases. In particular, fixed costs are overestimated for large multi-plant firms because they are complement to the omitted firms' incentive of restraining entry to reduce cannibalization. On the contrary, fixed costs will be underestimated for smaller multi-plant firms because they are contaminated by the omitted incentive of boosting entry to crowd out competitors. In general, the estimated parameters assuming separated entry do not perform well in predicting the "true" multi-plant firms' decisions, and subsequently could generate inaccurate aggregated outcomes if multi-plant firms are large and plant interdependency is strong.

Canonical trade models that studying oligopolists, such as BEJK and Atkeson and Burstein (2008a), do not clearly distinguish a plant and a firm.[4] This is concerning given mounting evidences that support differences between the two economic entities (i.e., Rossi-Hansberg et al., 2018; Hsieh and Rossi-Hansberg, 2019; Aghion et al., 2019; and Cao et al., 2017), especially when multi-plant firms are prevalent in oligopolistic industries. My model addresses the omission and offers testable implications on prices, markups and market shares across firms with different bundles of plants. As an extension of BEJK and the subsequent work by Holmes et al. (2011, 2014), and De Blas and Russ (2015), the distributions of costs and markups derived from the multi-plant firm model nest those in single-plant setting.[5] The model, therefore, yields more generalized insights on firm-level decisions regardless of single- or multi-plant owners.

I also build on and contribute to a vibrant area of ongoing research that explores interdependencies in firm-level decisions. In the international trade literature that explores firms' extensive margin with spillovers, this paper shares the most similarities with Tintelnot (2017), Antras et al. (2017), and Alfaro et al. (2021). An obviously important application of my model is multinational firms. Similar to this paper that studies substitutabilities in firms' multi-plant production, Tintelnot (2017) focuses on multinational production facing positive export platform sales and cannibalization. Nevertheless, his work evaluates all possibilities in a very small location set and the method is not easily scalable. I overcome the challenge by combining theoretical properties from the submodular game with a solution algorithm for combinatorial discrete choice problem. Not about multinational firms per se, Antras et al. (2017) and Alfaro et al. (2021) speak to complementarities in firms' global sourcing decisions and firms' export participation dynamics, respectively. In estimation, Antras et al. (2017) and this paper both employ a three-step procedure to separately identify demand and supply primitives. Alfaro et al. (2021) add time dimension to the combinatorial choices of export destinations. However, none of these papers adopt a game setting with multiple players. They all model a monopolistic competitive market and treat firms as infinitesimal with constant markups, whereas I highlight a small group of sizable firms competing oligopolistically and exploiting geographical advantages to increase markups. This key difference makes my

[4]BEJK take different view of the world compared to the way Atkeson and Burstein (2008a) modeling oligopoly in trade. In Atkeson and Burstein (2008a), each firm produces a distinct good in a specific sector and firms maximize profits given imperfect substitution within a sector and across sectors. In contrast, BEJK model multiple producers producing the same good and there are a continuum of imperfectly substituted goods. I follow BEJK by assuming firms produce a homogeneous good. I acknowledge that this assumption may limit the scope of industries where the framework can be applied. However, an advantage for adopting BEJK is that it requires less firm-level data than Atkeson and Burstein (2008a).

[5]Bernard et al. (2003) have markup distribution being impervious to any characteristics of market structure. Subsequent papers by Holmes, De Blas and others generalize the model to incorporate the effects of finite number of firms in a market. My model is closer to the later development that recognizes the granularity of firms.

model more suitable for analyzing policy questions in industries that are dominated by a few large firms.

This paper joins a literature in the field of industrial organization that analyzes how retailers set up distribution networks in space, such as Jia (2008) and Holmes (2011). The technique to solve for combinatorial discrete choice is developed based on Jia (2008). Jia (2008) imposes a "supermodular" condition on the return function, meaning that there is only positive spillover among chain stores. Eckert et al. (2017) generalize Jia (2008) and use the repetitive fixed points search algorithm to deal with either "supermodular" or "submodular" problem. The algorithm is further developed by Hu and Shi (2019) to a continuum of monopolistically competitive firms over a monotonic type space. The recent paper by Oberfield et al. (2020) handles the combinatorial optimization in an entirely different way. They derive a limit case where the optimal plant density is differentiable in a continuous space. On the other hand, an alternative approach to combinatorial problem is using moment inequality to partially identify parameters without solving the model, such as Holmes (2011). The downside for this method is the difficulty to perform counterfactual experiments.

Other than interdependency in location choices, the paper also broadly relates to multi-product firms when there is cannibalization among products within the same firm. Specifically, the endogenous markup distributions drawn from this paper and Nocke and Schutz (2018) share the same shape.

The remainder of the paper is structured as follows. In Section 2, I lay out the model and propositions derived from it. In Section 3, I start by describing the data generating process from the multi-plant firm model, and then propose a method to consistently estimate it using commonly found data in the literature. An approximation using single-plant firm assumptions is presented and compared to the "truth" in Section 4. The paper ends with conclusions and a list of possible applications in Section 5.

2 A Model of Multi-plant Firms

This section sets out a theory of production locations, export and prices for multi-plant firms with market power. A firm and a plant are distinct, albeit related, economic entities. A plant can potentially serve the demand locally and elsewhere. A firm internalizes cannibalization within itself and competition with rivals by deciding where to produce, and how much each of its plants should charge. For simplicity, all that differentiates a firm from other potential entrants prior to entry are fixed costs of building plants.[6] Once fixed costs are paid, plants are differentiated by

[6]The ex-ante heterogeneity across potential entrants can be extended to include firm-level productivity differences, but they are omitted for simplicity and would require additional data to be identified. I incorporate this extension in

production and trade costs associated with their locations, and a stochastic term that indicates their productivity level. Each firm selects its optimal plant sites by maximizing total expected single-period profits. I consider a partial equilibrium environment by focusing on interdependent entry and price competition between oligopolies in an industry.

The model features a static simultaneous entry game with complete information. I identify the competition and cannibalization effects from the plants' spatial distribution pattern. This approach abstracts from a number of dynamic considerations. For example, it does not allow for preemptive entry (Igami and Yang, 2013; Zheng, 2016) nor does it allow for any learning process by firms (Arkolakis et al., 2018). One may also raise the concern about firms' additional considerations in a dynamic setting, such as how sunk costs and negative scrap values can deter relocation of a plant.[7] In this regards, "relocation" decisions under a dynamic model are evaluated differently compared to "location" decisions under a static model. Since the main goal of this model is to study the interdependency in multi-plant firms' location decisions in a steady state and to compare long run equilibria under different policy regimes, incorporating transition dynamics is beyond the scope of this paper. Empirically, given the difficulty of solving the CDC problem, it is also extremely computational intensive to extend the framework to a dynamic setting unless imposing additional assumptions.

Formally, there is a finite number of discrete geographical units, $m \in \mathcal{M}$. A firm, $f \in \mathcal{F}$, chooses a subset of locations $\mathcal{L}_f \subseteq \mathcal{M}$ to set up plants, where a plant is indexed by $\ell \in \mathcal{L}_f$.[8] The firm owns $N_f = |\mathcal{L}_f|$ number of plants.

I start with the description of demand and then turn to the problem of multi-plant firms.

2.1 Demand

Demand is characterized for a single product bought by a continuum of consumers $i \in \mathcal{D}_m$ on a unit interval in m. The aggregated local demand is Q_m units of the good. I assume an isoelastic demand at the location level, given by

$$Q_m = A_m P_m^{-\eta}, \tag{1}$$

where $-\eta < -1$ is the price elasticity of demand to be consistent with profit maximization of monopolists.[9] The local price index of the good is P_m, and the exogenous demand shifter is A_m. I

Appendix B.1.

[7] Scrap values can be positive or negative depending on whichever is higher, proceeds from selling the land or payments to cleanup.

[8] I assume a firm cannot have more than one plant at a location. Essentially, a firm choosing a set of plants is equivalent to choosing a set of locations to produce.

[9] A special case of isoelastic demand is CES preferences.

formulate local demand instead of demand of each consumer for reasons that will be clear later. To give a quick preview here, firms only obtain knowledge on how consumers may differ after plants are built. At the earlier stage when a firm still chooses plant locations, it expects to charge the same price for all consumers in m. Therefore, to construct a firm's expected profits, I only need the demand information at the location level.[10]

2.2 The multi-plant firm's problem

A multi-plant firm decides where to establish production operations and how to serve consumers at every location. A finite number of oligopolists appoint their plants to produce the same product facing the aforementioned demand function. The timing of game is that at $t = 1$, firms simultaneously decide the set of locations to build plants in order to maximize expected profits, and pay the respective fixed costs. The set $\mathcal{F}$ of potential entrants is given. At $t = 2$, firms learn about the realized productivity of plants and decide to which consumer plants supply. Plants compete in price. For simplicity, I assume there is no fixed cost of exporting and every plant can be potential supplier of each consumer at every location.[11] I solve the model by backward induction.

2.2.1 Production decisions given plant locations

Each location $m \in \mathcal{M}$ is characterized by an exogenous productivity level T_m, as well as local equilibrium characteristics that firms take as given, namely, the demand shifter A_m and costs of input w_m. Inputs to produce the good are immobile across locations. Trade between any two locations bears an iceberg trade cost. For example, firm f has a plant in ℓ. The cost of transporting the good from ℓ to a consumer at the center of m is $\tau_{\ell m}$.

Conditional on firm f produces in a set $\mathcal{L}_f$ of locations, for each location $\ell \in \mathcal{L}_f$, the firm converts one bundle of inputs into a quantity $Z_{f\ell i}$ of the good for consumer $i \in \mathcal{D}_m$ at constant return to scale. The term $Z_{f\ell i}$ represents idiosyncratic shock specific to a plant-consumer pair. Examples of such factors include some degree of relationship specificity and internal distance between consumers in m towards the center. Rather than dealing with each $Z_{f\ell i}$ separately, I assume that they are realizations of independently and identically distributed random draws from a Fréchet distribution. The cumulative distribution function of the productivity that firm f's plant

[10] One can easily add more structure to the demand side, such as CES of the good and substitutes at consumer level and then aggregate to a location. However, additional demand parameters add no benefit in solving the firm's problem but further complicate the calibration of the model.

[11] Fixed costs of exporting at firm level could be incorporated, as in Tintelnot (2017), but they are omitted for simplicity and would require additional data to be identified. However, if the fixed costs of exporting are associated with the set of plant locations, then a firm would no longer select the minimum cost plant to serve a destination consumer, and the model would lose tractability.

in ℓ is

$$F_\ell^{draw}(z) = \Pr[Z_{f\ell i} \le z] = \exp(-T_\ell z^{-\theta}).$$

Dispersion of productivity is represented by θ. The bigger θ is, the more similar are the productivity draws.

Combining productivity, input and trade costs, the marginal cost of supplying the good from a plant in ℓ to consumer i in m is therefore

$$C_{f\ell im} = \frac{w_\ell \tau_{\ell m}}{Z_{f\ell i}}, \forall \ell \in \mathcal{L}_f, i \in \mathcal{D}_m. \tag{2}$$

It is distributed as

$$F_{\ell m}^c(c) = \Pr[C_{f\ell im} \le c] = 1 - \exp\left(-\phi_{\ell m} c^\theta\right),$$

where $\phi_{\ell m} = T_\ell (w_\ell \tau_{\ell m})^{-\theta}$ indicates the capability of location ℓ serving location m.

A caveat here is that plants at the same location are ex-ante identical regardless of ownership. The setup is analogous to Antras et al. (2017) where as long as the productivity draws are from a location-specific distribution, any firm specific term is absorbed. One may argue to include a firm's core productivity parameter to shift its plants' productivity as in Tintelnot (2017) such that more productive firm will build more productive plants on average. As I demonstrate in Appendix B.1, it is straightforward to incorporate additional firm-level heterogeneity into the benchmark model. However, estimation of the model becomes substantially more data-hungry.[12] Although firms are not endowed with core productivities, we will show later that a firm having more plants at efficient (higher T) locations implies a more productive firm overall. Therefore, the ex-ante heterogeneity across firms is fully loaded on a firm's number of plants and their locations, i.e. the extensive margin.

Plants engage in Bertrand competition in a nested structure. Every consumer in a location is served by its lowest-cost supplier. If firms are single-plant, the winning firm is constrained not to charge more than the second-lowest marginal cost, the standard setting in BEJK. In the case of multi-plant firms, the firm headquarters decide prices of all their plants instead of plant managers. The firm headquarter will internalize competition between its own plants and coordinate their pricing. As a result, the winning plant will not undercut its *sister* plants owned by the same firm, until the next lowest-cost plant is owned by a competitor. The price charged is limited by the marginal cost of the lowest-cost plant owned by the second-lowest cost firm. Instead of fully characterize cost ranking across all plants, what really matter are the lowest-cost plant within a firm and the two lowest-cost firms.

[12] To estimate the set of firm core productivity parameters, I would need each firm's market share in every location which is not commonly available. I welcome researchers who have the relevant data to use the extended version of the model in the Appendix.

First, define the kth lowest cost among firm f's plants serving consumer i in m as $C_{k,fi(m)}$. The distribution of the lowest marginal cost can be easily derived as

$$F^c_{1,fm}(c) = \Pr[C_{1,fi(m)} \leq c] = 1 - \exp(-\Phi_{fm} c^\theta), \tag{3}$$

where $\Phi_{fm} = \sum_{\ell \in \mathcal{L}_f} \phi_{\ell m}$ refers to the capability of a firm f serving location m. The assumption of Fréchet distributed productivities becomes handy in the derivation due to its grounding in the extreme value theory. If a firm selects the best available technology from a distribution, its productivity will also follow an extreme value distribution. While technical advantages dictate this choice, empirical distributions of productivity are typically bell-shaped in the literature, which also in favor of the Fréchet specification.[13]

Based on equation (3), the firm-level expected marginal cost to consumers in m is

$$E\big[C_{1,fi(m)}\big] = \Gamma\left(\frac{\theta+1}{\theta}\right) \Phi_{fm}^{-\frac{1}{\theta}}. \tag{4}$$

More plants at favorable (high $\phi_{\ell m}$) locations necessarily lower the firm's effective marginal cost.[14] Intuitively, one more production location grants the firm an additional cost draw, and greater competition among plants. Although plants within the same firm coordinate their pricing strategies, they still fight for being the ultimate supplier. Intense internal competition reduces the expected marginal cost at the firm level. More plants also imply that the average shipping distance to consumers is shorter and thus additional savings on trade costs for firms. Moreover, the effect of an additional plant is larger when it is located somewhere cheaper to produce and closer to consumers. The properties of the minimum cost distribution for a multi-plant firm allow me to establish the following result (the proof is straightforward and omitted in the main text).

Proposition 1: An additional production location to the firm's active location set strictly decreases its lowest cost of supplying the good to all consumers in expectation.

Second, define $C_{1,i(m)}$ and $C_{2,i(m)}$ be the lowest and second-lowest marginal cost across all firms for consumer i in m. Conditional on sales originating from firm f's plant at location ℓ, $C_{1,i(m)} \equiv C_{f\ell i(m)}$ and $C_{2,i(m)} \equiv \min_{g \neq f, g \in \mathcal{F}}\{C_{1,gi(m)}\}$. I show in Appendix A.1 that the condi-

[13] Another commonly used distribution is Pareto. Although Fréchet and Pareto distribution both have fat right tails, they are very different on the left side. The former density is bell-shaped whereas the latter density is downward-sloping throughout. Other candidate could be log normal that is very hard to be distinguished from Fréchet, or truncated distributions. However, properties in truncated distributions are more obscured and hard to applied to my model without strong support of empirical evidences. A thorough examination and comparison of distributions can be found in Head (2011) and Kotz and Nadarajah (2000).

[14] The implication is in contrast to Oberfield et al. (2020) in which they focus on the span-of-control cost and more plants will reduce a firm's efficiency. Nevertheless, we both have that favorable locations reduce the marginal costs of plants and firms.

tional joint distribution of the lowest and second-lowest firm-level cost of supplying the good to a consumer at m is

$$F^c_{12,m|f}(c_1, c_2) = 1 - e^{-\Phi_m c_1^\theta} - \frac{\Phi_m}{\Phi_{fm}} \left(1 - e^{-\Phi_{fm} c_1^\theta}\right) e^{-\left(\Phi_m - \Phi_{fm}\right) c_2^\theta}, \tag{5}$$

for $c_1 \leq c_2$, where $\Phi_m = \sum_{f \in \mathcal{F}} \sum_{\ell \in \mathcal{L}_f} \phi_{\ell m}$ denotes the sourcing potential of location m over all plants. Notice that the conditional joint distribution is independent of which plant in firm f sells to consumers in m. The rationale is that the cost ranking between plants within a firm does not matter to the game except for determining the firm-level minimum cost.

When the number of firms approaches to infinity, the limit distribution of equation (5) is what BEJK use for the joint distribution of two lowest costs. When firms are finite but single-plant, equation (5) takes the form of the joint distribution in Holmes et al. (2011). Therefore, the two key features that differentiate my model from previous literature are granularity and multi-plant production. We would expect that the additional dimensions on extensive margin based on BEJK type of model bring richer implications on how markups vary across firms.

I now turn to describe the price and markup distributions in this model. The competition structure implies a partially limit pricing strategy, where the lowest-cost plant charges a minimum between the monopoly price and the lowest marginal cost of its head-to-head competitors. Mathematically, the price charged to consumer i in m is $P_{i(m)} = \min\{\bar{\mu} C_{1,i(m)}, C_{2,i(m)}\}$, where the monopoly markup $\bar{\mu} = \eta/(\eta - 1)$.

Conditional on sourcing from firm f, the firm decides its winning plant to charge consumers in location m at a price following the distribution,

$$F^p_{m|f}(p) = F^c_{12,m|f}(p, p) + \frac{\Phi_m}{\Phi_{fm}} \left(1 - e^{-\Phi_{fm} \bar{\mu}^{-\theta} p^\theta}\right), \tag{6}$$

with derivation shown in Appendix A.2. A closer look at equation (6) reveals that the first term comes from the cost ladder, while the second term is derived from the probability to charge the monopoly price. Pass-through is zero if the limit pricing prevails, and one otherwise. Combining with Proposition 1, a firm with larger and favorably located plant set lowers its average price. The expected price charged by firm f to consumers in m is

$$E\big[P_{fm}\big] = \Gamma\left(\frac{\theta + 1}{\theta}\right) \frac{\Phi_m}{\Phi_{fm}} \left(\left(\Phi_m - (1 - \bar{\mu}^{-\theta})\Phi_{fm}\right)^{-1/\theta} - \left(\Phi_m - \Phi_{fm}\right) \Phi_m^{-\frac{\theta+1}{\theta}}\right), \tag{7}$$

Closely related, firm f's markup in location m is realization of a random draw from a shifted

Pareto distribution truncated at the monopoly level,

$$F^{\mu}_{m|f}(\mu) = \begin{cases} 1 - \frac{1}{(1-s_{fm})\mu^{\theta}+s_{fm}} & 1 \leq \mu < \bar{\mu} \\ 1 & \mu \geq \bar{\mu} \end{cases}, \tag{8}$$

where $s_{fm} = \Phi_{fm}/\Phi_m$ indicates the relative competitiveness of firm f to its rivals. It is the sole shifter of the markup distribution. See Appendix A.3 for derivation. Specifically, for $1 \leq \mu < \bar{\mu}$, a firm owning more plants in favorable locations charges higher markup. For $\mu \geq \bar{\mu}$, I compute the probability of firm f charging monopoly markup given the second-lowest cost equals to

$$\frac{1 - e^{-\Phi_{fm}(\bar{\mu}/c_2)^{-\theta}}}{1 - e^{-\Phi_{fm}c_2^{\theta}}}.$$

Therefore, knowing that the otherwise price charged will be c_2, the firm is more likely to exploit the maximum markup if it is more capable of supplying the product through plant allocation because the efficiency gap to its next lower cost rival is wider. We observed that more dispersed plants indicated by smaller θ also increase the likelihood of charging the monopoly price.

The markup distribution again generalizes what is in single-plant firm models. In the case of infinite number of firms competing head-to-head, the markup distribution converges to equation (11) in BEJK. I summarize results in the following proposition.

Proposition 2: Holding the competitors fixed, (i) an additional production location to the firm's active location set weakly decreases its average price charged to all consumers; (ii) an additional production location to the firm's active location set weakly increases its average markup charged to all consumers.

Lastly, I explore the implications of our model for the bilateral trade volume across locations, aggregating from firms' decisions. With firms' cost distributions in equation (3), the probability that firm f supplies to a consumer in m is

$$s_{fm} = \int_0^{\infty} \prod_{g \neq f. g \in \mathcal{F}} \left(1 - F^c_{1,gm}(c)\right) dF^c_{1,fm}(c) = \frac{\Phi_{fm}}{\Phi_m}. \tag{9}$$

Essentially, the probability equals to a firm's relative competitiveness of supplying the good compared to all other head-to-head competitors. Since all consumers are uniformly distributed on a unit interval, the probability of supplying to a consumer is the same as the expected fraction of consumers captured in m.

Proposition 3: An additional production location to the firm's active location set strictly increases the share of consumers sourcing from it, holding the competitors fixed.

Similarly, suppose a set $\mathcal{F}_\ell$ of firms produce at ℓ and $N_\ell = |\mathcal{F}_\ell|$, the probability that location ℓ exports to a consumer in m is

$$s_{\ell m} = \int_0^\infty \prod_{k \neq \ell, k \in \mathcal{M}} \left(1 - F^c_{1,km}(c)\right) dF^c_{1,\ell m}(c) = \frac{N_\ell \phi_{\ell m}}{\Phi_m}, \tag{10}$$

where $F^c_{1,\ell m}(c) = 1 - \exp\left(-N_\ell \phi_{\ell m} c^\theta\right)$ characterizes the distribution of the lowest-cost plant at ℓ across all firms entered. The probability represents location ℓ's competitive advantage. The more plants, the higher local efficiency, the lower input costs and the lower trade costs in a location, the larger share m sourcing from it. Different from BEJK which do not have a measure of firms, this paper shows more firms producing in a location is pro-competitive despite firms being symmetric for simplicity.

Recall that $\phi_{\ell m} = T_\ell (w_\ell \tau_{\ell m})^{-\theta}$, equation (10) can be transformed to resemble a standard gravity equation. The trade elasticity is shaped by the Fréchet parameter θ as in Eaton and Kortum (2002).

The model propositions derived from this section are summarized and represented visually in Figure 1.

2.2.2 Choice of plant locations

A firm chooses the set of plant locations from a finite discrete space $\mathcal{M}$ to maximize the expected total profit summing over its plants. To complete the expected total profit function, I first present the expected variable profit, with details presented in Appendix A.4 and A.5.

$$E\big[\pi_f\big] = \kappa \sum_m A_m \left(\bar{R}_{fm} - \bar{C}_{fm}\right), \tag{11}$$

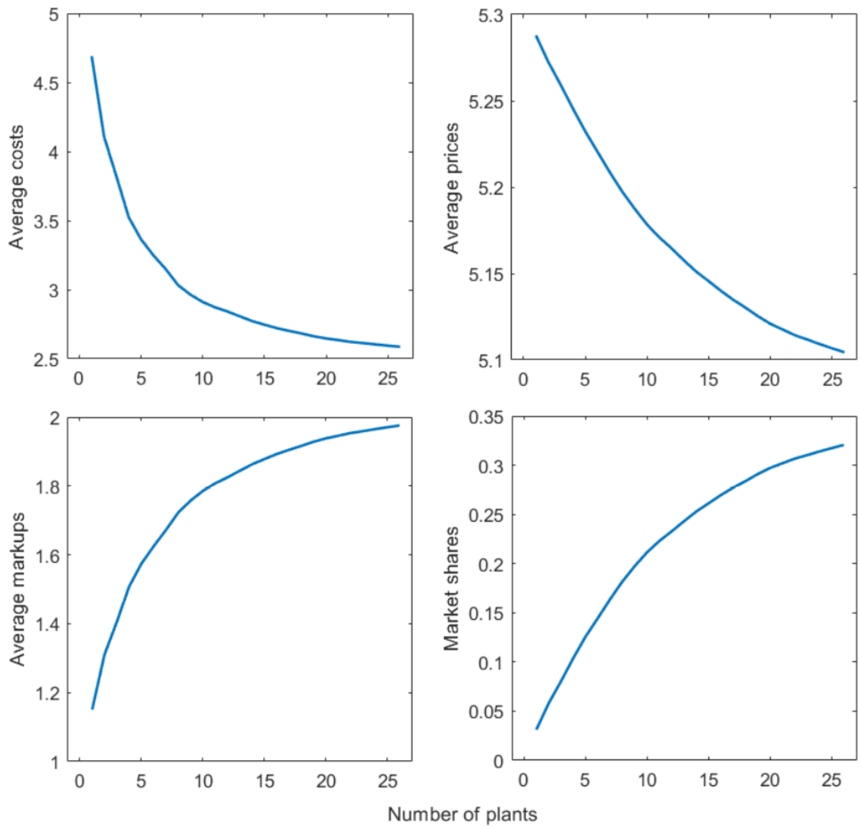
where the constant $\kappa = \Gamma\left(\frac{\theta+1-\eta}{\theta}\right)$, and

$$\bar{R}_{fm} = \left(\Phi_m - (1-\bar{\mu}^{-\theta})\Phi_{fm}\right)^{-\frac{1-\eta}{\theta}} - \left(\Phi_m - \Phi_{fm}\right)\Phi_m^{-\frac{\theta+1-\eta}{\theta}},$$

$$\bar{C}_{fm} = \Phi_{fm} \times \Bigg[(\theta+1-\eta)(\Phi_m - \Phi_{fm}) \int_1^{\bar{\mu}} \mu^{-\theta-2} \left(\Phi_m - (1-\mu^{-\theta})\Phi_{fm}\right)^{-\frac{2\theta+1-\eta}{\theta}} d\mu + \bar{\mu}^{-\theta-1}\left(\Phi_m - (1-\bar{\mu}^{-\theta})\Phi_{fm}\right)^{-\frac{\theta+1-\eta}{\theta}}\Bigg].$$

The expectation of variable profit is taken over random productivity draws for all plant-consumer pairs. It depends on the capability of supplying the good from all of the firm's plants and its

Figure 1: Firm-level changes with number of plants



Notes: The figure displays a firm's expected marginal costs of supplying the good, expected prices, markups, and the fraction of consumers served, averaging across locations, with respect to the number of plants it owns. Parameter values are the same as the Monte Carlo simulation in Section 3.1. The firm of interest here is firm B, taking firm M and small firms as given. The curve is drawn by adding firm B's plant one by one, starting from the location with the highest profitability, until reaching its optimal plant set of 26 plants.

competitors'. More importantly, each plant is not separately additive. Cannibalization effects defines multi-plant firms' location problem to be combinatorial optimization.

In order to have a well defined expected variable profit, I must restrict $\eta - 1 < \theta$. The same restriction has been seen in the literature, such as Eaton and Kortum (2002), Eaton et al. (2011), and Bernard et al. (2003), with η representing demand elasticity and θ being the heterogeneity of suppliers in production. The condition ensures that suppliers are competitive enough such that consumption is not concentrated on a few of them. Mathematically, the condition is in need for a well-behaved κ after taking the expectation over a function of Fréchet distributed stochastic term, knowing that Gamma function is sensitive to parameter values at the negative support. One may contrast the condition to what is in Antras et al. (2017) that guarantees the supermodularity of sourcing decisions. A clear difference is mentioned in Antras et al. (2017) footnote 11 where θ in their setting is no longer heterogeneity of final good suppliers but input producers although η

still remains to be demand elasticity among final varieties. The submodularity or supermodularity property of equation (12) will be explained in more details in Section 2.3.1.

Although the restriction on η and θ has little to do with submodularity of the profit function, discussing the comparative statics of a firm's profit with respect to these two parameters help us to understand the firm's optimal plant location strategy. Figure 1 implies that a firm obtains positive marginal payoff by adding one more plant to its existing active set. However, when θ is high meaning that plants are more homogeneous, the lowest cost among a firm's own plants will not reduce much after building another plant. Furthermore, when demand is less elastic (low η), the firm's variable profit responses weaker to cost reductions and gains less.

A multi-plant firm incurs plant-specific fixed costs for setting them up, $\{FC_{f\ell}, \forall \ell \in \mathcal{L}_f\}$.[15] Fixing the same set of plant locations, firms would expect exactly same variable profits because plants at the same location are symmetric. Therefore, what drives one firm to have more plants than the other is having lower fixed costs on average. Location choices are affected by the firm's idiosyncratic fixed costs at different locations and profitability given competitors' fixed costs and location choices. A firm thus solves

$$\max_{\mathcal{L}_f \subseteq \mathcal{M}} E\big[\Pi_f(\mathcal{L}_f)\big] = E\big[\pi_f(\mathcal{L}_f)\big] - \sum_{\ell \in \mathcal{L}_f} FC_{f\ell}. \tag{12}$$

Finally, I close the model with the local price index, which is a composite of prices that all firms charge to consumers in m.

$$\begin{aligned} P_m &= \sum_{f \in \mathcal{F}} E[P_{m|f}] \times s_{fm} \\ &= \Gamma\left(\frac{\theta+1}{\theta}\right) \Phi_m^{-1/\theta} \times \left[(1-N) + \sum_{f \in \mathcal{F}} \big(1 - (1-\bar{\mu}^{-\theta}) s_{fm}\big)^{-1/\theta}\right], \end{aligned} \tag{13}$$

where $N = |\mathcal{F}|$ is the given number of firms. The equation explains how variation on local prices is channeled through plants spatial distribution globally.

2.3 Equilibrium

So far, I have yet discussed in detail how to find the equilibrium of the interdependent production locations in this game-theoretic model. I will first show the existence of equilibrium depending on

[15] There is a strand of literature concerning greenfield entry versus merger and acquisition. However, the case of M&A need not to be fundamentally different from my benchmark model. Acquisition price can be seen as the fixed cost, except the case that the acquisition price depends on the seller's residual value that is past dependent. If then, fixed costs are also endogenous and need to be solved using a dynamic model.

important properties of firms' profit functions. Then, I discuss how to address the issue of multiple equilibria like many other simultaneous entry games with complete information.

2.3.1 Existence of pure-strategy Nash equilibrium

In a single-agent location problem, with $|\mathcal{M}|$ number of potential production locations, a firm selects among $2^{|\mathcal{M}|}$ possible configurations. Theoretically, one can use the brute force approach to calculate firm profits for all combinations of locations and pick the set yielding the maximum profit. However, the computational cost grows exponentially when $|\mathcal{M}|$ gets large. In general, it is also not guaranteed that the optimal location set is unique for a discrete choice problem even in the case of single player. So what is sufficient condition to ensure a global maximum and how to find the optimal in a cost-efficient way? Fortunately, Eckert et al. (2017) provide a solution where if the objective function exhibits single crossing differences, we can iteratively and repetitively refine the combinatorial discrete choice set and the process always converges to a unique equilibrium.

Noticing that equation (12) is submodular in a multi-plant firm's own strategy, meaning that the marginal value on total profit of adding location ℓ by firm f is decreasing in the number of other locations that f entered. Specifically, from the propositions, the variable profit of a firm increases from expanding its plant location set, but the marginal gain diminishes with more plants due to self-cannibalization. Submodularity in the firm's profit function is a sufficient condition for single crossing differences, suggesting that unprofitable locations will remain to be unprofitable when enlarging the set and profitable ones will remain to be profitable when shrinking the set. Leveraging the monotonicity, Eckert et al. (2017) generalize the method that is first developed in Jia (2008) to the case of submodular profit functions and further show that we can always reach a unique maximizing vector by partitioning the lattice and repetitively applying the algorithm. I will discuss how to implement the algorithm with more details in Section 3.2.3.

In a multi-agent location game, existence of equilibrium, and in particular a pure strategy Nash equilibrium (PSNE), becomes much more challenging. There are three aspects of complexity in the game described in my multi-plant firm model: discrete choices as firms decide to enter or not, multidimensional as each strategy is defined as a vector of ones and zeros, and strategic substitutes as players face competition. Attributed by the first two points, for a two-player $|\mathcal{M}|$-location game, the domain of strategies is an enormous set of $2^{2|\mathcal{M}|}$ number of configurations. Although the third point seems to be prevalent in many applications, tackling games that exhibit strategic substitutes is not as straightforward. Past literature has shown that there are substantial imbalances in existence and characterization of equilibrium between games with strategic substitutes and strategic complements (i.e. Vives, 1999; Jackson and Zenou, 2015; Jensen, 2005). An advantage of studying a game with strategic complements is that a PSNE always exists following Tarski's fixed point theorem (Tarski et al., 1955) and Topkis' monotonicity theorem (Topkis, 1978). In such case, the equilibrium set is a complete lattice and highly structured where players benefit from coordination and typically the greatest PSNE is also Pareto optimal (Milgrom and Roberts, 1990; Zhou, 1994). However, existence of PSNE is not generally true in games with strategic substitutes.[16]

The multi-plant firm model in Section 2 is a submodular game. With multidimensional strategies, submodular games are games in which the marginal returns to any component of the player's strategy drop with increases in other components of the player himself and the competitors' strategies. I have demonstrated above that the profit function exhibits decreasing differences in a firm's own strategy due to self-cannibalization. The same holds for the marginal profit to be decreasing in the firms' joint strategy space due to competition. This model does not yield forces that could make plants being strategic complements to each other.[17] Neither does it imply any admissible parameter setting that leads to supermodularity of the payoff function. The most relevant papers in recent development of proving existence of a PSNE in submodular games are Dubey et al. (2006) and Jensen (2010). They restrict attention to aggregative games in which the payoff of a player only depends on his own strategy and an aggregate of others' strategies (or what's called "quasi-aggregative games" in Jensen (2010) when the strategy set is multidimensional). In our context, the firm's profit, equation (12), is a function of its own location strategy $\mathcal{L}_f$ and a weighted additive aggregate of rivals' locations, Φ_m.[18] Submodularity implies that the game contains strategic substitutes. Moreover, if the plant location strategy sets are compact and the profit function is upper semi-continuous, using best-reply potential game properties, Jensen (2010) proves that a quasi-aggregative game of strategic substitutes has a PSNE.[19] [20] I will illustrate how to find the equilibrium in Section 3.2.3 using a duopoly game, although theoretically it can be extended to more than two players.

Proposition 4: For a $|\mathcal{F}|$-player, $|\mathcal{M}|$-location game in Section 2 with profit function exhibiting submodularity for all players, the set of pure-strategy Nash equilibria is not empty.

[16] One can refer to Example 1 in Jensen (2005) where no equilibrium exists for a strategic substitutes game.

[17] Introducing agglomeration forces could easily make the firm's profit function intractable.

[18] Mapping to the notation in Jensen (2010), the aggregator $g(\mathcal{L}) = \sum_{f\in\mathcal{F}} \Phi_{fm}(\mathcal{L}_f)$. The interaction functions are $\sigma_f(\mathcal{L}_{-f}) = \sum_{g\neq f, g\in\mathcal{F}} \Phi_{gm}(\mathcal{L}_g)$. The shift-functions $F_f(\sigma_f(\mathcal{L}_{-f}), \mathcal{L}_f) = \sigma_f(\mathcal{L}_{-f}) + \Phi_{fm}(\mathcal{L}_f) = g(\mathcal{L})$. Therefore, our multi-plant firm model is a quasi-aggregative game by Definition 1 in Jensen (2010).

[19] According to Corollary 1 in Jensen (2010), the quasi-aggregative game has to satisfy Assumption 1 and 2 for a PSNE to exist. Assumption 1 is satisfied because the location game presented here features strategic substitutes and therefore every firm's best-reply correspondence is a decreasing selection. Assumption 2 is also satisfied through a monotonic transformation of the shift-functions.

[20] Eckert et al. (2017) impose additively separable condition to a player's profit function to prove existence of PSNE in a game exhibiting single-crossing differences, meaning that the profit function is additively separated to a player f specific part and a common part of all players' actions. This is a much stronger sufficient condition than what is needed in Jensen (2010).

2.3.2 Multiple equilibria

A common concern in estimating discrete games is the existence of multiple equilibria. The fact that for a given set of parameters and covariates, there may be more than one equilibrium outcome, raises the well-known coherency problem in econometric inference (Heckman, 1978; Tamer, 2003). In the absence of interdependency across locations, for a $2\times2\times1$ (two players choosing enter one location or not) game with competition, the Nash equilibria is that either firm enters and the other stays out. With interdependency, the game would accommodate more equilibria.

Surveying the literature, there are four main approaches to deal with the multiplicity of equilibria.[21] The first is to model the probabilities of aggregated outcomes that are robust to multiplicity. For example, in the simplest $2\times2\times1$ game, the number of entrants is unique although the firm identity is undetermined (Bresnahan and Reiss, 1990; Bresnahan and Reiss, 1991; Berry, 1992). However, information on firm heterogeneity is lost. Should I use it in this paper, I would not be able to estimate the fixed costs distributions which are firm-location specific.

Second, one can embrace the multiplicity and take a bounds approach (Ciliberto and Tamer, 2009; Pakes et al., 2015). The method sets identify parameters which could be too large to be informative. Lack of point identification becomes difficult when performing counterfactual exercises. Estimating a bound also causes inference to be computationally intensive, such as placing confidence region on the set.

The third approach to multiple equilibria—the one taken here—is to choose an equilibrium by imposing certain entry sequence. I simulate the data and estimate the model by selecting the equilibrium that would be chosen by the low cost firm. While I model the entry game as static, the assumption is convenient in avoiding multiple equilibria.[22] In principle, estimates could be sensitive to the equilibrium selected and the predetermined order of entry. I assume the "true" equilibrium is selected because the focus of the exercise is to investigate whether assuming separated entry is a good approximation to interdependent plant location model. Introducing additional uncertainty complicates the comparison.

More recent development of the literature is around specifying a more general equilibrium selection rule that is a function of covariates and observables, as in Bajari et al., 2010. The solution requires computing all equilibria and an equilibrium selection parameter as part of the primitives to be estimated together with the model. Although this approach is more general than imposing certain sequence of entry, the computational burden to calculate all equilibria in an interdependent entry game is too high.

[21] Ellickson and Misra (2011) provide a thorough discussion on estimating static discrete games, especially methods for dealing with the issue of multiple equilibria.

[22] The same approach has been taken by Jia (2008), Atkeson and Burstein (2008b), Eaton et al. (2012), Edmond et al. (2015) among many others.

3 Structural Analysis

In this section, I lay out an estimation procedure of the multi-plant firm model with commonly available data. Instead of using the real data set, I generate data based on the multi-plant firm model and then present results to show consistency of the estimates to the true DGP.

3.1 DGP in Monte Carlo simulations

I consider a region consisting of 50 locations with one single-plant local firm in each of the location.[23] Each location independently draws logarithm of the demand shifters A, local productivity T, and input costs w, all from a uniform distribution on the open interval $(4, 6)$. As for trade costs, I set $\tau_{\ell m} = (1 + t) \times d_{\ell m}^{0.2}$, where t is the 10% tariff for trade flow between any location dyads, and distance $d_{\ell m}$ is specified as one plus the absolute difference between any two index ℓ and m from 1 to 50.

There are two large entrants deciding where to produce out of the 50 potential locations given the group of small incumbents. Productivities of plants are distributed Fréchet with shape parameter $\theta = 5$. Price elasticity of demand is set to be $-\eta = -2$. Firm B (for "big") and firm M (for "medium") draw plant-level fixed costs from a log-normal distribution,

$$\begin{pmatrix} \log(FC_{B\ell}) \\ \log(FC_{M\ell}) \end{pmatrix} \sim N \left[\boldsymbol{\mu}^F = \begin{pmatrix} 1.5 \\ 2 \end{pmatrix}, \;\; \boldsymbol{\Sigma}^F = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix} \right].$$

Firm B on average is cheaper to build plants than firm M. Given that the exercise only have two firms, I let them have perfectly correlated draws with common standard deviation, suggesting same perception on fixed costs of establishing across locations. Should I allow for imperfect correlation across firms, there need to be more firms to identify such parameter.

Based on these "true" primitives of the multi-plant firm model, I simulate the model 100 times to generate 100 sets of data. Averaging across all the simulations, firm B has 26 plants and firm M has 13 plants consistent with firm B facing lower fixed costs.[24] Firm B accounts for around 32% of the total expenditure on the product over the 50 locations, and the share for firm M is 21%. Among consumers served by the two firms, the prices paid to firm M is on average 0.98 times of the prices paid to firm B, and firm M faces marginal costs 1.1 times of those for firm B. Hence, firm M's average markup is lower. Since the datasets are generated from the multi-plant firm model, it is not surprising to see that these comparisons between the two firms corroborate the propositions. Aggregating firms' exports to location level, the average domestic market share

[23] For simplicity, I assume a group of small firms to make each location a potential candidate for production, i.e. finite input costs. Therefore, all locations can potentially sell to one another.

[24] The average out of 100 simulations is rounded to the nearest integer.

is 21% corresponding to the share of product consumed in a market that is produced locally.

3.2 Multi-plant firm estimation

The typical dataset that econometricians observe involves a combination of aggregated data at location level and limited firm level data. Other micro data, such as prices or shipping flow for individual plant, is not necessarily available to researchers. I hereby propose a procedure to estimate the full model with minimal data requirement, namely bilateral trade volume between locations $Q_{\ell m}$, trade costs $\tau_{\ell m}$, demand shifters A_m, and each firm's plant locations $\{\mathcal{L}_f, \forall f =$ B, M, single-plant firms$\}$. For simplicity, I assume iceberg trade costs are known truthfully, and so do the demand shifters.[25] The key parameters of interest in estimation are then narrowed down to $\{\theta, \eta, \boldsymbol{\mu}^F, \boldsymbol{\Sigma}^F\}$.

The estimation is performed in three steps. First, I use a gravity-type regression to estimate the composite of locations' production capability $T_\ell w_\ell^{-\theta}$ and the Fréchet dispersion θ. The sourcing probability derived from the model provides a natural link between theoretical implication and the bilateral trade data. Next, I project local consumption on the model-consistent price index constructed using the estimates from last step, and estimate the demand elasticity $-\eta$ through nonlinear least squares. What has been obtained in the first two steps is critical for constructing firms' expected profit as a function of plant location configurations and fixed costs. In the final step, I match the predicted optimal plant locations to the actual ones to pin down parameters that govern the fixed cost distribution via method of simulated moments. Separability in estimation allows me to reduce dimensionality of the problem and save computational cost. More importantly, I can verify that the profit function is well defined before implementing the combinatorial optimization algorithm in the last step.

3.2.1 Step 1: Estimation of location production capability and plant productivity dispersion

The first step is to estimate each location's production capability summarized by the term $T_\ell w_\ell^{-\theta}$, and the dispersion of plants' productivities θ. To do so, I take the plant locations as given and exploit differences in trade originated from other local attributes, such as productivity, input costs, and trade costs. Recall that equation (10) provides the probability of m sourcing from ℓ. Empirically, the model-predicted sourcing probability is associated with the trade share in volume, i.e.

[25] In practice, econometricians observe distance between locations and tariff rates instead of the ad valorem trade costs. They need to additionally estimate the trade cost elasticity with respect to distance. Similarly, econometricians observe location characteristics and need to estimate A as a function of the covariates. However, I assume these elasticities are known at true values here because they are not central to firms' trade-offs in this paper.

$s_{\ell m} = \frac{Q_{\ell m}}{Q_m}$. Transform equation (10) to its estimable version,

$$\frac{Q_{\ell m}}{Q_m} = \exp\left[\text{FE}_\ell + \text{FE}_m - \theta\tau_{\ell m} + \epsilon_{\ell m}\right], \tag{14}$$

where the origin fixed effect $\text{FE}_\ell = \ln\left(N_\ell T_\ell w_\ell^{-\theta}\right)$, and the destination fixed effect $\text{FE}_m = -\ln\Phi_m$. I estimate the gravity regression via Poisson Pseudo Maximum Likelihood (PPML) due to the consistency it delivers under general conditions and its capability of incorporating zeros as clearly explained in Silva and Tenreyro (2006) and Head and Mayer (2014).

Estimating equation (14), I obtain the Fréchet shape parameter directly from the trade elasticity. As for the component $T_\ell w_\ell^{-\theta}$ for each location, I need to tease out the number of plants N_ℓ from the origin fixed effects. The model presumes that local efficiency and input costs are underlying economic conditions without general equilibrium feedback of plants spatial distribution on factor markets. Consequently, I can substitute N_ℓ with the observed data on plant locations.[26] Another issue is that the location production capabilities can only be estimated up to a scale. Normalization affects the absolute level of variable profits and subsequently fixed costs. I will explain how to recover the scale parameter in step 2.

After the first step, I obtain estimates of an efficiency and input costs combination for each location and the key parameter θ. Effects of these components on firm profits depend on the price elasticity, $-\eta$. I now turn to estimating the demand.

3.2.2 Step 2: Estimation of demand

To estimate demand featured in equation (1), I combine it with the local price index derived from the model. Recall in equation (13), the price index is a function of the estimated location production capability and Fréchet parameter $\hat{\theta}$, data on trade costs and plant locations, and the unknown price elasticity. Denote the estimated location production capability from step 1 as $\ln\widetilde{T_\ell w_\ell^{-\theta}} = \ln\widehat{T_\ell w_\ell^{-\theta}} + \lambda$, where λ is the normalization parameter and $\widehat{T_\ell w_\ell^{-\theta}}$ are the absolute levels used in DGP. Then, $\ln P_m(\widehat{T_\ell w_\ell^{-\theta}}) = \ln P_m(\widetilde{T_\ell w_\ell^{-\theta}}) + \lambda/\hat{\theta}$. Assuming demand shifter A_m is observed, I estimate the following using nonlinear least squares because the price elasticity affects demand also through firm markups,

$$\ln\frac{Q_m}{A_m} = \kappa_2 - \eta\ln P_m(\eta; \widetilde{T_\ell w_\ell^{-\theta}}), \tag{15}$$

[26]In DGP, I assume all locations have at least one plant. Therefore, I arbitrarily eliminate the potential issue of zero plant and unidentified local characteristics. In real data, this problem would restrict our potential location set to only those with observed positive production, causing selection bias.

where the constant $\kappa_2 = -\eta\lambda/\hat{\theta}$. Hence, not only the price elasticity is estimated, but also I adjust the level of local production capability to be used later for constructing comparable firm profits as in DGP.

In practice, one should also find valid instruments for prices and perform the above via generalized method of moments, even for the constructed price index because the endogenous plant locations are also used to form the price index. Some candidates for price instruments could be cost shifters in nearby locations. In this Monte Carlo simulation, for simplicity, I assume away the endogeneity problem so that all the estimation error can be traced to the last step of the estimation.[27] I can then focus on examining how well the model-predicted plant locations match the simulated data, and keep a clear comparison between the estimates under the multi-plant versus the single-plant firm assumption.

3.2.3 Step 3: Estimation of fixed costs

Having the necessary elements for constructing firms' expected payoff, the last step is to solve for firms' optimal location sets and estimate the fixed costs of establishing plants. The model does not yield a closed form solution to firms' location choices conditional on the location observables and estimated vector of parameter values. Hence, I would need simulation methods to get the simulated estimate of entry probability for each firm-location pair.[28] I assume the fixed costs are log-normally distributed as in DGP with a common shape but different mean. I then draw a 2×50-dimensional matrix of $FC_{f\ell}$ for a large number of times.[29] For each repetition, the firm maximizes total expected profits by choosing where to build the set of plants taken rivals' locations as given.

In a duopoly game with 50 potential locations, firms select from 2^{100} which is a magnitude of 10^{30} number of configurations. The model requires $(\eta - 1)/\theta < 1$. Estimates from the first two steps would verify whether the condition is satisfied. In this submodular game, marginal profit of adding a plant is strictly decreasing with the size of the active plant set. A firm would always enter a location which subtracting it from the full set will bring negative marginal profit, and vice versa, he would always stay out of a location which adding it to the null set brings negative marginal

[27]Note that the trade costs and demand shifters are observed at true values, so the constructed price index in the second step will perfectly match the simulated "true" price if η is at its true value. Using instruments generated from DGP in estimation will introduce additional error.

[28]There is actually another level of simulation for firm markups. Notice that the expected variable profit function (12) involves numerical integration over the markup. I use a stratified random sampling method in order to obtain good coverage of the higher markup. I define intervals from 1 to $\bar{\mu} = \eta/(\eta - 1)$, [1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.95, 1.97, 1.99, $\bar{\mu}$]. I then draw 5 uniform random numbers within these intervals. The draws receive a weight inversely proportional to the length of the interval. The integral part of the profit function is approximated by $\int_1^{\bar{\mu}} f(\mu) \approx \sum_{s=1}^{S} w_s f(\mu_s)$.

[29]For the fixed cost draws, I follow Antras et al. (2017) to use quasi-random numbers from a van der Corput sequence which have better coverage properties than usual pseudo-random draws. I use 300 simulation draws for estimation.

profit. Define firm f's marginal profit of including ℓ in a location strategy $\mathcal{L}_f$ as

$$\Delta^\ell \Pi_f(\mathcal{L}_f) = \Pi_f(\mathcal{L}_f \cup \ell) - \Pi_f(\mathcal{L}_f \setminus \ell).$$

Specifically, the AE-Repetitive algorithm in Eckert et al. (2017) works as follows. In the single-player case, starting from $\mathcal{L}_f = \mathcal{M}$ which contains all locations, $\ell \in \mathcal{L}_f^1$ if $\Delta^\ell \Pi_f(\mathcal{M}) > 0$. Also at the other extreme, starting from $\mathcal{L}_f = \emptyset$ which contains no entries, $\ell \notin \mathcal{L}_f^1$ if $\Delta^\ell \Pi_f(\emptyset) < 0$. The first round of mapping confirms some of the coordinates. Now iterate the mapping until a complete equilibrium location set is reached with no more refinement can be made. When there are indefinite coordinates, Eckert et al. (2017) prove that the set of possible vectors can be sliced to any two subsets and then operate the mapping on each of the subsets separately. Slicing and mapping is repeatedly done until a unique optimal location vector $\mathcal{L}_f^*$ emerges.

In the duopoly case, there will be multiple equilibria as explained in Section 2.3.2. I assume the larger firm, firm B, enters first which leads to the equilibrium most profitable for firm B. The same ordering is used in DGP. Therefore, firm B starts to solve for his best response using the algorithm by taking firm M enters nowhere. Then firm M finds his best response given firm B's initial strategy. Two firms take turn to react until both best responses converge. Since the duopoly game is submodular, the iteration reaches a PSNE, $\{\mathcal{L}_B^*, \mathcal{L}_M^*\}$. The speed of convergence in a game with best-response potential properties is exponential derived in Swenson and Kar (2017).[30]

I identify the fixed costs of building plants via method of simulated moments (MSM). With any guess of $\{\boldsymbol{\mu}^F, \boldsymbol{\Sigma}^F\}$, I can solve for the equilibrium plant locations for the two firms using the above-mentioned algorithm. Due to the small firm sample, I fix perfectly correlated location draws between the two firms as in DGP. Essentially, parameters of interest are $\mu_B^F, \mu_M^F, \sigma^F$. Moments are to match the model-predicted and the observed values of (a) number of firm B's plants; (b) number of firm M's plants; (c) difference between the average production capability for locations that firm B produces and those that firm B stays out; (d) difference between the average production capability for locations that firm M produces and those that firm M stays out. The first two moments are informative about the mean of fixed costs for the two firms. The last two moments help to pin down the dispersion of fixed costs distribution. The smaller the dispersion, the less entry decisions vary by fixed costs but more by local profitability. Thus, firms produce where the location production capability is higher relative to others.

Formally, the vector of moment functions, $g(\cdot) \in \mathbf{R}^4$, specifies the differences between the observed equilibrium outcomes and those predicted by the model. The following moment condition

[30] In Table A.4, I present examples of two firms choosing among different location sets. The average time of convergence for 10 potential locations is around 0.09 seconds averaging 1000 simulations. It takes a maximum of three rounds of iteration to find the best response for two firms.

is assumed to hold at the true parameter value $\delta_0 = \{\boldsymbol{\mu}^F, \boldsymbol{\Sigma}^F\} \in \mathbf{R}^3$,

$$E\big[g(\delta_0)\big] = 0. \tag{16}$$

A MSM estimator is such that

$$\hat{\delta} = \arg\min_{\delta} \frac{1}{|\mathcal{M}|}\Bigg[\sum_{\ell=1}^{|\mathcal{M}|}\hat{g}(\delta)\Bigg]' W \Bigg[\sum_{\ell=1}^{|\mathcal{M}|}\hat{g}(\delta)\Bigg], \tag{17}$$

where $\hat{g}(\cdot)$ is the simulated estimate of the true moment function and W is a weighting matrix.[31] I use the identity matrix and weight the moments equally in the following estimation of the simulated data because precision of the estimates is not a first-order concern when comparing results from the interdependent entry model to those assuming separated entry.

The complexity when having spatial correlation is that the moment functions $g(\cdot)$ are no longer independent across locations. In order for the method of moments estimators using dependent cross section data to be consistent, Conley (1999) proves the sufficient condition is that the dependence among observations should die away quickly as the distance increases. In the current model setup, competition between plants becomes weaker when locations are further apart due to trade costs.

3.3 Results from multi-plant firm estimation

Table 1 presents results of the three-step estimation procedure. It should be expected that the first two steps estimate Fréchet shape parameter θ and price elasticity η perfectly because DGP and estimating regressions are based on exactly same set of equations in the model. There is not measurement error introduced either since I substitute the "true" data on demand shifters, trade cost and number of plants when regressing (14) and (15). The estimated location production capabilities from fixed effects in gravity regression are also perfectly correlated with the simulated data on $T_\ell w_\ell^{-\theta}$.

Let's focus on the estimated parameters governing the fixed cost distribution. The estimated mean of the fixed costs distribution for firm B and firm M are close to the true parameters. They are also estimated with high level of precision indicated by the small average standard error and mean absolute deviation. I also compute the simulated coverage probabilities of the confidence interval, and found them remarkably close to the nominal coverage probability: for the 95% confidence sets the simulated coverage probabilities are 97% and 94% respectively. The results support that

[31] The discrete choice decisions makes the objective function non-smooth and the firm's problem not globally convex. The shortcoming is that I cannot guarantee the solution I find is the global optimum of the problem. To address this issue, I tried the particle swarm optimization algorithm to search through 100 starting points. All sets of starting points resulted in close outcomes.

average fixed costs of firm B and firm M are well identified by matching the number of plants each firm owns. So is the difference in average fixed cost of building a cement plant by the two firms. The estimated standard deviation σ^F is slightly upward biased, but the true parameter value is still less than one standard error away from the estimate. The below 95% simulated coverage probability indicates that the dispersion of fixed cost distribution is not as strongly identified as the mean. The sample of 50 locations and two firms can make the inference suffer from asymptotic size distortion besides lack of independence.

Comparing the predicted plant entry probabilities for the two firms with the local profitability represented by the production capability across locations, Figure 2 clearly illustrates a positive correlation between the two. More importantly, the figure shows what's left to be explained in entry decisions is driven by variation in fixed costs across locations. The dominant role of fixed costs is particularly striking for those with fewer plants, causing some locations with relative production advantage do not appear to be attractive to firms.

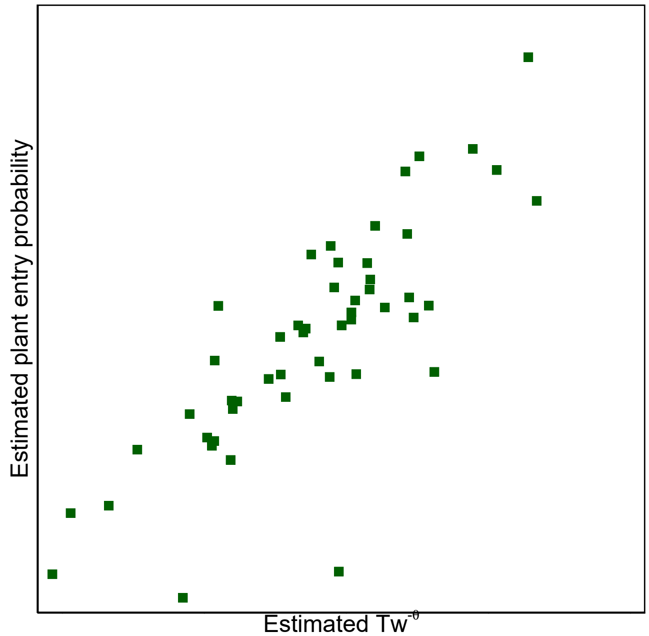
Table 1: Estimated Parameters

	(1) True values	(2) Avg. estimates	(3) Avg. std. err.	(4) MAD	(5) Simulated coverage
θ	5	5.0	0	0	1
η	2	2.0	0	0	1
μ_B^F	1.5	1.500	0.216	0.092	0.97
μ_M^F	2	2.155	0.286	0.087	0.94
σ^F	0.5	0.693	0.235	0.105	0.83

Based on estimates and estimated standard errors from 100 simulated datasets, the table reports average of the estimates, average of the asymptotic standard error, and mean absolute deviation (MAD) of the estimates. Simulated coverage probabilities of the 95% (nominal coverage) confidence intervals are constructed as the following. First, compute the confidence interval for each dataset using estimate $\pm$ 1.96 $\times$ standard error. Second, count the proportion of datasets for which the confidence interval contains the value of the true parameters.

I examine the estimation results by describing the model's fit of the data in Table 2. For the moments that were targeted to match in estimation, namely number of plants built by each firm, the prediction almost exactly fits the data. I also report the correlations between the predicted and observed number of firm B plants and firm M plants in each location, which is 0.64 and 0.60 respectively. These correlations are not targeted in estimation, but still indicate a reasonable fit. Since number of plants affect trade flow between locations, I further check the correlation between predicted share of trade volume for each location dyads and the observed shares used in the first-step gravity regression. They are highly correlated at 96.5%. The fraction of consumption supplied

Figure 2: Predicted entry probabilities and production capabilities by location



by domestic production is also matched pretty well. Lastly, I use Figure 3 to compare the predicted consumption and production across locations against the simulated data. Both dimensions distribute tightly along the 45-degree line and explain more than 90% of the data. Overall, multiple test results establish confidence in the proposed multi-plant firm estimation method accounting for interdependent entry decisions.

4 Single-plant Approximation

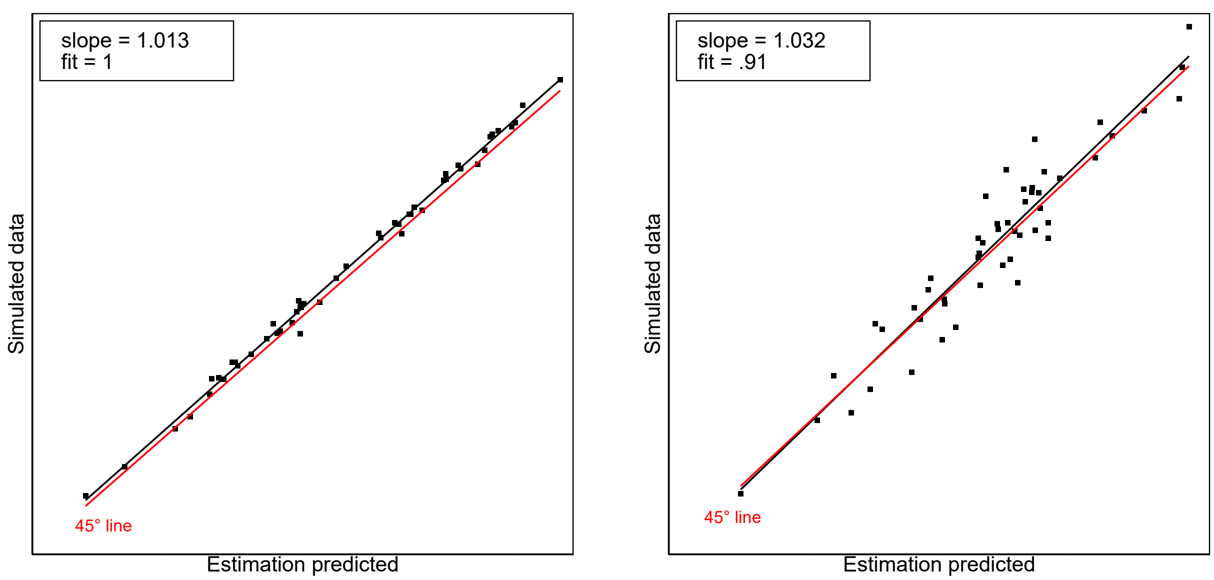
Having described all the mechanisms in a multi-plant firm framework, I proceed to address a pressing question when applying this model: does interdependent entry matter? Realistically, a multi-plant firm could operate in a continuum degree of control over its plants, with one extreme being complete oversight of all its production locations and the other being full delegation to local managers. The latter is equivalent to treating each establishment as a single-plant firm. The framework in this paper highlights two distinct features in modeling decisions of multi-plant firms (MP) compared to single-plant firms (SP). One is that multi-plant firms coordinate prices among plants owned. Price setting at the firm level instead of plant gives the firm incentive to expand plant set to crowd out rival firms. The other is that multi-plant firms internalize cannibalization through strategically placing the set of plants. A firm has incentive to reduce its number of plants or locate them further apart to minimize self-cannibalization. These two counteracting forces are absent in

Table 2: Goodness of Fit

	Data	Model predicted
Targeted moments		
Firm B number of plants	25.7	25.7
Firm M number of plants	13.2	13.2
Non-targeted moments		
Domestic market share	0.210	0.216
Corr of trade shares		0.965
Corr (firm B entry prob, sample locations)		0.641
Corr (firm M entry prob, sample locations)		0.600

Statistics for simulated data and model prediction are averages across 100 repetitions, and thus number of plants is not integer. Correlations are calculated by stacking the respective variables from each repetition.

Figure 3: Fit for consumption and production



(a) Consumption (average over 100 simulations)

(b) Production (average over 100 simulations) a world with only single-plant firms. They compete head-to-head with all other plants and enter if the expected profit net of fixed costs of oneself is positive.

Although imposing single-plant assumption is inconsistent with the multi-plant firm's objective to maximize total profits of all its plants through both pricing and entry decisions, it is an empirically handy approach for researchers especially when studying discrete choice decisions because one does not need to solve combinatorial optimization. Therefore, in this section, I focus on measuring the difference in entry decisions by ignoring interdependent entry in estimation for a multi-plant firm DGP. Specifically, the model is preserved with spillovers from price coordination and cannibalization, but instead of solving a set of plant locations, I estimate the cost parameters assuming separate entry across locations. One can think of this alternative as a case where a local manager decides whether to build a plant or not given all other plants' locations. The profit he/she expects is distributed by the parent firm based on a fraction of the firm's total profit.

Under the same model, now let's compute the expected profit for firm f at location ℓ. Because pricing strategy is coordinated at the firm level, the price distribution for every plant owned by the same firm is identical as shown in equation (6). Hence, the expected variable profit for a particular plant is proportional to the firm's profit based on the share of consumers sourced from this plant over all firm f's consumers. With the Fréchet distributed productivities, the sourcing probability of consumers in m from plant ℓ over the set of firm f's plants is

$$s_{f\ell m} = \frac{\phi_{\ell m}}{\Phi_{fm}}. \tag{18}$$

Multiplying the probability by equation (11) and subtracting the associated fixed costs, I obtain the expected profit of a plant owned by firm f at ℓ is

$$E\big[\Pi_{f\ell}\big] = E\big[\pi_{f\ell}\big] - FC_{f\ell}, \tag{19}$$

where $E\big[\pi_{f\ell}\big] = \kappa \sum_m s_{f\ell m} A_m \left(\bar{R}_{fm} - \bar{C}_{fm}\right)$.

Maintaining the same parametric assumption that fixed costs follow a log normal distribution shifted by firm ownership, and then taking logs of the plant entry condition $E\big[\pi_{f\ell}\big] > FC_{f\ell}$, the empirical form of plant entry probability under single-plant approximation is

$$\Pr\left[\ell \in \mathcal{L}_f\right] = \Phi\left(\frac{1}{\sigma^F}\ln E\big[\pi_{f\ell}\big] - \frac{\mu_f^F}{\sigma^F}\right). \tag{20}$$

Since the focus is to compare separate versus interdependent entry, estimating equations from the first two steps in the multi-plant firm procedure remain the same. Aggregated outcomes at location level are not affected by the change of entry assumption if keeping the observed number of plants the same. Therefore, I can use same estimates from the first two steps to calculate plant-level profit. Inverting the coefficient of the expected profit at plant $f\ell$ gives a direct estimate of the standard deviation of log fixed costs, σ^F. I estimate equation (20) as a binary probit with the constructed $\ln E\left[\pi_{f\ell}\right]$ on the right-hand side, together with a firm dummy.[32]

4.1 Monte Carlo results

Table 3 summarizes and compares the estimation results under SP separate entry assumption and MP interdependent entry assumption. I compute the average fixed costs based on the mean equation of the log-normal distribution, $\exp(\mu + \sigma^2/2)$. The larger firm B's fixed cost tends to be overestimated when neglecting interdependency, while firm M's fixed cost is slightly lower. With SP approximation, the simulated coverage probabilities of confidence interval for the estimated distribution mean are substantially lower than the 95% benchmark, 47% and 78% respectively. This result demonstrates clear bias in estimating fixed costs without considering that plant locations are interdependent on one another. The standard deviation of log fixed costs is unexpectedly a close and precise estimate to the true value. However, the under-performance in coverage probability of σ^F in the case of MP is not as salient as the poor coverage of μ^F in the case of SP.

To understand why fixed costs are estimated to be higher in single-plant approximation for firms with more plants, one has to link back to the two externalities that are neglected for single plants, coordination and cannibalization. Recall that I have demonstrated in Section 2.3.1 that marginal profit of a multi-plant firm is positive with the number of plants as the firm is in better position to compete against rivals. There are two countervailing forces against expanding the plant set: cannibalization effects and additional fixed costs. These two push the size of plant set to the same direction. Moreover, cannibalization intensifies with the number of plants a firm owned. Hence, a large firm cares more about competition among its own plants and build fewer relative to the counterfactual SP scenario. The single-plant approximated fixed costs need to be higher to match the observed plants number from a multi-plant firm DGP. On the contrary, for a multi-plant firm that is not as big, the concern for cannibalization is weaker and biases on fixed costs drop with no constraint that the biases are necessarily upward. I leave finding the cutoff in plants number that balances the MP and SP results for future research.

Suppose I use the overestimated fixed costs of firm B from SP approximation to predict its entry decisions, Figure 4 presents that firm B would have entered all locations with significantly smaller probabilities, averaging across 100 simulated datasets. Firm M would also have entered less aggressively on average, but by a smaller difference. Across simulations, panel (a) in Figure 5 is a visual illustration of the fixed costs results in Table 3. In panel (b), predicted number of plants

[32] I cluster the standard error at location level to account for correlated draws. Coefficients for each set of simulated data are significant at 1%.

Table 3: Single-plant Approximated Parameters

	True values	Single-plant		Multi-plant	
	(1)	(2) Avg. estimates	(3) Coverage	(4) Avg. estimates	(5) Coverage
μ_B^F	1.5	1.698	0.47	1.500	0.97
μ_M^F	2	2.210	0.78	2.155	0.94
σ^F	0.5	0.534	1	0.693	0.83

Based on 100 simulated datasets, the table reports average of the estimates and simulated coverage probabilities of the 95% confidence interval. In column (2) and (3), for each dataset, I compute the parameters of interest transformed from the Probit coefficients estimated using equation (20), and their corresponding standard error using Delta Method (refer to Appendix C.1). Confidence intervals are then constructed. Next, I count the proportion of datasets for which the confidence interval contains the value of the true parameters. Column (4) and (5) are from Table 1.

for the two firms are matched exactly to the observables using multi-plant firm estimated fixed costs, but under-predicted using single-plant approximation. Researchers would predict 4 plants fewer (roughly 15%) for firm B and 2 plants fewer (roughly 13%) for firm M.

Figure 4: Entry probability across locations

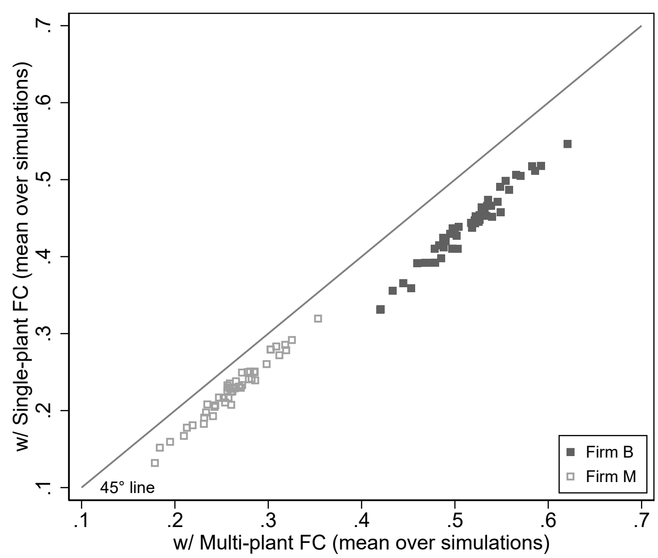
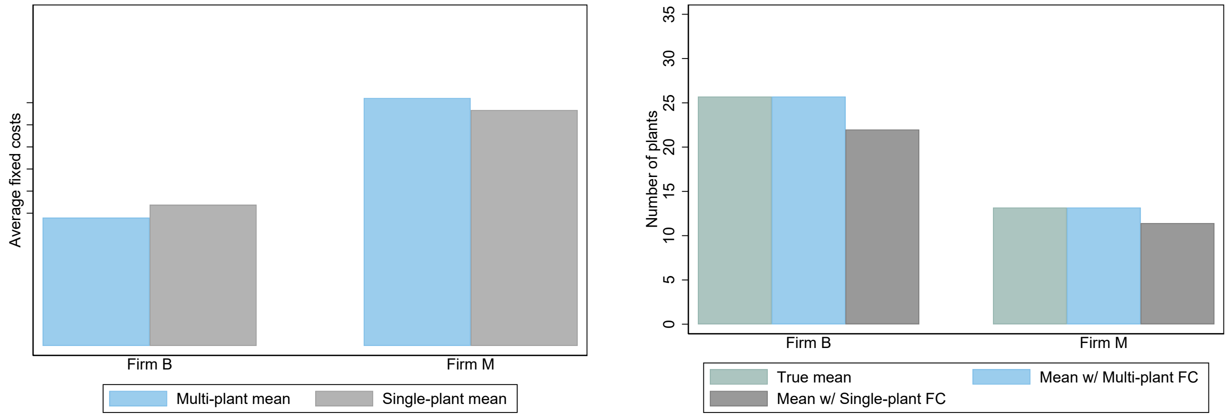


Figure 5: Firm-level comparison using multi- v.s. single-plant fixed costs across simulations

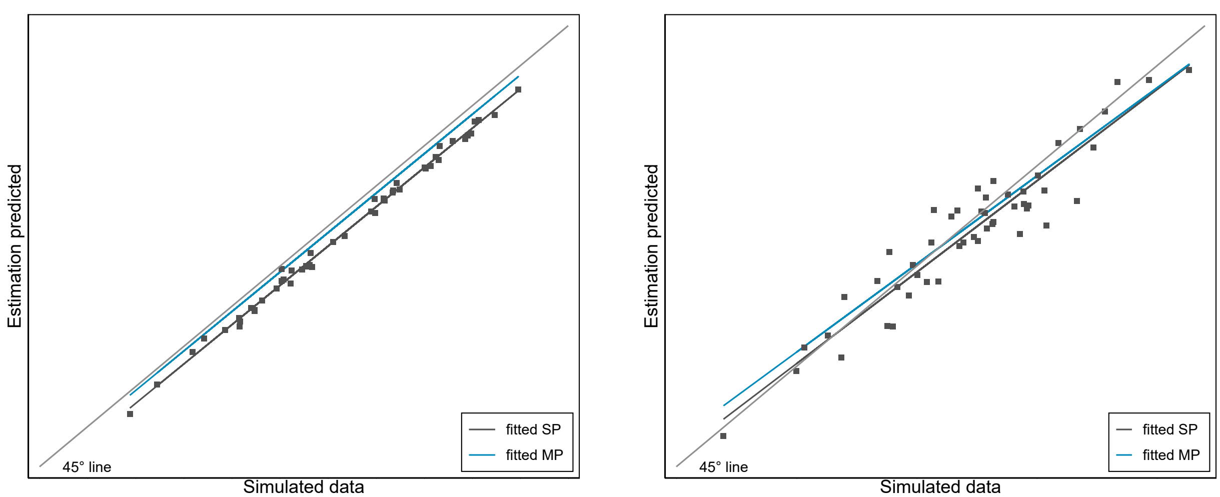


(a) Fixed costs (mean over locations)

(b) Number of plants (sum over locations)

Although the SP approximation underperforms at the firm level, the prediction on market aggregates does not appear to be deviated by a significant amount, if anything, slightly lower than the simulated consumption and production across locations shown by Figure 6. The reason is that consumption or production depends on plants spatial distribution as well as location characteristics such as its production costs and efficiency, and accessibility to nearby locations. These other factors remain to be consistently estimated under the single-plant assumption, and thus alleviate the biases in the predicted plant locations. The three-step estimation procedure is appealing because separability ensures some model primitives are intact from the firm entry assumptions.

Figure 6: Consumption and production using multi- v.s. single-plant fixed costs across locations



(a) Consumption (mean over simulations)

(b) Production (mean over simulations)

There are two things that could make the single-plant approximation deteriorate. The first is that firms are large and own many plants. The second is that the cannibalization effect is strong indicated by a higher θ. Both would strengthen a multi-plant firm's motive to build fewer plants and further differentiate the firm's decisions from a single plant.

The takeaway from this exercise is that assuming separated entry decisions for multi-plant firms generates estimation biases which are heterogeneous across firms depending on their sizes. In particular, it has trouble approximating plant locations of large firms. Since large multi-plant firms typically control a dominant share of economic activities in an industry, failure to account for interdependent sets of plants may misguide us from the correct outcomes.

5 Conclusions

In this paper, I develop a multi-plant oligopoly model with endogenous and interdependent location decisions. On the intensive margin, the model derives multi-plant firms' pricing and markup rule in closed forms that generalize findings for single-plant firms in BEJK and others. It also characterizes the extensive margin of multi-plant production and quantitatively solves a firm's optimal set of plant locations. The two margins interact in a way such that a set of properties emerge. More and favorably located plants increase the production advantage of a firm, improve its capability to compete against rivals, and enhance the firm's market power to charge higher markup. At the same time, there are two forces against the expansion of plant set: cannibalization and fixed costs. Marginal payoff of opening up new plants diminishes because of business stealing among a firm's own plants. Hence, a firm strategically places an additional plant to the right location until the marginal payoff cannot cover its fixed costs. The framework points to the importance of spatial interdependency raised from positive and negative spillovers among plants within the same firm.

A key contribution of this paper is to overcome the empirical challenge of solving combinatorial discrete choice problem for a multi-player game. Under a parametric restriction on demand elasticity and competition intensity among plants, the game is submodular, meaning that all plants are substitutes. For a firm's own plants, competition dominates the positive spillover from the firm's increased market power. Submodularity and aggregative property of the location game underpin the methodology to solve for combinatorial optimization by eliminating many location configurations. I extend the algorithm in Eckert et al. (2017) for a game-theoretic framework and show that we can always find a pure strategy Nash equilibrium. Moreover, the model generates a gravity trade equation that can be leveraged to separately identify model primitives, which further reduces the computational cost of estimating the full model.

The paper goes on to use simulated evidences to show that if the multi-plant firm model is the truth, assuming separated entry could make poor predictions if firms are large and cannibalization effects are strong. Estimates from a binary Probit are biased, and the direction of biases depends on firm sizes.

This machinery of analysing multi-plant firms can be applied to evaluate a wide range of policies. Specifically, how firms respond to policy changes by reorganizing plants and subsequent effects on prices, market structure, and welfare. For example, it is a known issue that carbon tax would induce leakage through firms moving to pollution haven. Other shocks, such as trade war or mergers of competitors, could also generate unintended effects through firm-level responses. Since multi-plant firms are prevalent in many industries and control a large share of production and trade, studying policy effects without a careful treatment of multi-plant production is worrisome. One caveat is that application of the model is restricted to firms producing a homogeneous good. Products such as cement, steel or paperboard are likely to be suitable candidates. It also depends on at which level the product of interest is classified.

References

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2019). A theory of falling growth and rising rents. Technical report, National Bureau of Economic Research.

Alfaro, A., Morales, E., Manuel Castro Vincenzi, J., and Fanelli, S. (2021). Firm export dynamics in interdependent markets. Technical report, working paper.

Antras, P., Fort, T. C., and Tintelnot, F. (2017). The margins of global sourcing: Theory and evidence from us firms. *American Economic Review*, 107(9):2514–64.

Arkolakis, C., Papageorgiou, T., and Timoshenko, O. A. (2018). Firm learning and growth. *Review of Economic Dynamics*, 27:146–168.

Atkeson, A. and Burstein, A. (2008a). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031.

Atkeson, A. and Burstein, A. (2008b). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031.

Bajari, P., Hong, H., and Ryan, S. P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568.

Bernard, A. B., Eaton, J., Jensen, J. B., and Kortum, S. (2003). Plants and productivity in international trade. *American economic review*, 93(4):1268–1290.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, pages 889–917.

Bresnahan, T. F. and Reiss, P. C. (1990). Entry in monopoly market. *The Review of Economic Studies*, 57(4):531–553.

Bresnahan, T. F. and Reiss, P. C. (1991). Entry and competition in concentrated markets. *Journal of Political Economy*, 99(5):977–1009.

Cao, D., Hyatt, H. R., Mukoyama, T., and Sager, E. (2017). Firm growth through new establishments. *Available at SSRN 3361451*.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of econometrics*, 92(1):1–45.

De Blas, B. and Russ, K. N. (2015). Understanding markups in the open economy. *American Economic Journal: Macroeconomics*, 7(2):157–80.

Dubey, P., Haimanko, O., and Zapechelnyuk, A. (2006). Strategic complements and substitutes, and potential games. *Games and Economic Behavior*, 54(1):77–94.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An anatomy of international trade: Evidence from french firms. *Econometrica*, 79(5):1453–1498.

Eaton, J., Kortum, S. S., and Sotelo, S. (2012). International trade: Linking micro and macro. Technical report, National bureau of economic research.

Eckert, F., Arkolakis, C., et al. (2017). Combinatorial discrete choice. In *2017 Meeting Papers*, number 249. Society for Economic Dynamics.

Edmond, C., Midrigan, V., and Xu, D. Y. (2015). Competition, markups, and the gains from international trade. *American Economic Review*, 105(10):3183–3221.

Ellickson, P. B. and Misra, S. (2011). Structural workshop paper—estimating discrete games. *Marketing Science*, 30(6):997–1010.

Head, K. (2011). Skewed and extreme: Useful distributions for economic heterogeneity. Technical report, Working paper series, University of British Columbia.

Head, K. and Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics*, volume 4, pages 131–195. Elsevier.

Head, K. and Mayer, T. (2019). Brands in motion: How frictions shape multinational production. *American Economic Review*, 109(9):3073–3124.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46(4):931–959.

Helpman, E., Melitz, M. J., and Yeaple, S. R. (2004). Export versus fdi with heterogeneous firms. *American economic review*, 94(1):300–316.

Holmes, T. J. (2011). The diffusion of wal-mart and economies of density. *Econometrica*, 79(1):253–302.

Holmes, T. J., Hsu, W.-T., and Lee, S. (2011). Plants, productivity, and market size with head-to-head competition. *Manuscript University of Minnesota (23 June)*.

Holmes, T. J., Hsu, W.-T., and Lee, S. (2014). Allocative efficiency, mark-ups, and the welfare gains from trade. *Journal of International Economics*, 94(2):195–206.

Hsieh, C.-T. and Rossi-Hansberg, E. (2019). The industrial revolution in services. Technical report, National Bureau of Economic Research.

Hu, K. and Shi, R. (2019). Solving combinatorial discrete choice problems in heterogeneous agents models. Technical report, Technical report.

Igami, M. and Yang, N. (2013). Cannibalization and preemptive entry of multi-product firms. *Rand Journal of Economics*, 36:908–929.

Irarrazabal, A., Moxnes, A., and Opromolla, L. D. (2013). The margins of multinational production and the role of intrafirm trade. *Journal of Political Economy*, 121(1):74–126.

Jackson, M. O. and Zenou, Y. (2015). Games on networks. In *Handbook of game theory with economic applications*, volume 4, pages 95–163. Elsevier.

Jensen, M. K. (2005). Existence, comparative statics, and stability in games with strategic substitutes. *University of Birmingham*.

Jensen, M. K. (2010). Aggregative games and best-reply potentials. *Economic theory*, 43(1):45–66.

Jia, P. (2008). What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, 76(6):1263–1316.

Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. World Scientific.

Milgrom, P. and Roberts, J. (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society*, pages 1255–1277.

Monderer, D. and Shapley, L. S. (1996). Potential games. *Games and economic behavior*, 14(1):124–143.

Nocke, V. and Schutz, N. (2018). Multiproduct-firm oligopoly: An aggregative games approach. *Econometrica*, 86(2):523–557.

Oberfield, E., Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2020). Plants in space. Technical report, National Bureau of Economic Research.

Pakes, A., Porter, J., Ho, K., and Ishii, J. (2015). Moment inequalities and their application. *Econometrica*, 83(1):315–334.

Ramondo, N. and Rodríguez-Clare, A. (2013). Trade, multinational production, and the gains from openness. *Journal of Political Economy*, 121(2):273–322.

Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2018). Diverging trends in national and local concentration. Technical report, National Bureau of Economic Research.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Swenson, B. and Kar, S. (2017). On the exponential rate of convergence of fictitious play in potential games. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 275–279. IEEE.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165.

Tarski, A. et al. (1955). A lattice-theoretical fixpoint theorem and its applications. *Pacific journal of Mathematics*, 5(2):285–309.

Tintelnot, F. (2017). Global production with export platforms. *The Quarterly Journal of Economics*, 132(1):157–209.

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations research*, 26(2):305–321.

Vives, X. (1999). *Oligopoly pricing: old ideas and new tools*.

Voorneveld, M. (2000). Best-response potential games. *Economics letters*, 66(3):289–295.

Zheng, F. (2016). Spatial competition and preemptive entry in the discount retail industry. *Columbia Business School Research Paper*, (16-37).

Zhou, L. (1994). The set of nash equilibria of a supermodular game is a complete lattice. *Games and economic behavior*, 7(2):295–300.

Appendices

A Model details

For simplicity of derivation, I invert the marginal cost and derive the following based on a plant's cost-adjusted productivity

$$\tilde{Z}_{f\ell im} = \frac{Z_{f\ell i}}{w_\ell \tau_{\ell m}}. \tag{A-1}$$

A.1 Conditional joint distribution of the lowest two cost firms

Since the conditional joint distribution of the lowest two costs is the same as that of top two productivities, the joint distribution of the first and second highest cost-adjusted productivity to market m conditional on firm f^* from ℓ^* winning the consumer is

$$\begin{aligned} F_{12,m}(z_1, z_2; \ell^*, f^*) &= \Pr\left(\tilde{Z}_{1m}(i) \leq z_1, \tilde{Z}_{2m}(i) \leq z_2 \mid \tilde{Z}_{1m}(i) > V\right) \\ &= \Pr\left(\tilde{Z}_{1m}(i) \leq z_2 \mid \tilde{Z}_{1m}(i) > V\right) \\ &\quad + \Pr\left(z_2 \leq \tilde{Z}_{1m}(i) \leq z_1, \tilde{Z}_{2m}(i) \leq z_2 \mid \tilde{Z}_{1m}(i) > V\right), \end{aligned} \tag{A-2}$$

where $V \equiv \max\{\tilde{Z}_{2m}(i), S\}$ and $S \equiv \max_{\ell \in \mathcal{L}_{f^*}, \ell \neq \ell^*}\{\tilde{Z}_{f^*\ell m}(i)\}$, given $z_1 > z_2$.

The distribution of S is

$$F_m^S(s; \ell^*, f^*) = \Pr(S \leq s; \ell^*, f^*) = \exp\left(-\left(\Phi_{f^*m} - \phi_{\ell^*m}\right) s^{-\theta}\right).$$

And the distribution of V is

$$F_m^V(\nu; \ell^*, f^*) = \Pr(V \leq \nu; \ell^*, f^*) = \exp\left(-\left(\Phi_m - \phi_{\ell^*m}\right) \nu^{-\theta}\right).$$

The first part of equation (A-2) can be simplified as

$$\begin{aligned} \Pr\left(\tilde{Z}_{1m}(i) \leq z_2 \mid \tilde{Z}_{1m}(i) > V\right) &= \frac{\Pr\left(V < \tilde{Z}_{f^*\ell^*m}(i) \leq z_2\right)}{\mathbb{P}_{f^*\ell^*m}} \\ &= \frac{\Phi_m}{\phi_{\ell^*m}} \int_0^{z_2} \left[\tilde{F}_{\ell^*m}^{draw}(z_2) - \tilde{F}_{\ell^*m}^{draw}(V)\right] dF_m^V(V; \ell^*, f^*) \\ &= \exp\left(-\Phi_m z_2^{-\theta}\right). \end{aligned} \tag{A-3}$$

Next, the second part of equation (A-2) is equal to

$$\frac{\Pr\left(z_2 \le \tilde{Z}_{f^*\ell^*m}(i) \le z_1, \tilde{Z}_{2m}(i) \le z_2, \tilde{Z}_{f^*\ell^*m}(i) > V\right)}{\mathbb{P}_{f^*\ell^*m}}$$
$$= \frac{\Pr\left(z_2 \le \tilde{Z}_{f^*\ell^*m}(i) \le z_1, \tilde{Z}_{2m}(i) \le z_2, \tilde{Z}_{f^*\ell^*m}(i) > S\right)}{\mathbb{P}_{f^*\ell^*m}},$$

where the equality is by definition of $\tilde{Z}_{2m}(i)$. The numerator can be further simplified as

$$\begin{aligned}
&\Pr\left(z_2 \le \tilde{Z}_{f^*\ell^*m}(i) \le z_1, \tilde{Z}_{2m}(i) \le z_2, \tilde{Z}_{f^*\ell^*m}(i) > S\right) \\
&= \Pr\left(z_2 \le S \le \tilde{Z}_{f^*\ell^*m}(i) \le z_1, \tilde{Z}_{2m}(i) \le z_2\right) + \Pr\left(S \le z_2 \le \tilde{Z}_{f^*\ell^*m}(i) \le z_1, \tilde{Z}_{2m}(i) \le z_2\right) \\
&= \int_{z_2}^{z_1} \left[\tilde{F}^{draw}_{\ell^*m}(z_1) - \tilde{F}^{draw}_{\ell^*m}(S)\right] \prod_{f \neq f^*} \tilde{F}_{1,fm}(z_2) dF^S_m(S; \ell^*, f^*) \\
&\quad + \int_0^{z_2} \left[\tilde{F}^{draw}_{\ell^*m}(z_1) - \tilde{F}^{draw}_{\ell^*m}(z_2)\right] \prod_{f \neq f^*} \tilde{F}_{1,fm}(z_2) dF^S_m(S; \ell^*, f^*) \\
&= \frac{\phi_{\ell^*m}}{\Phi_{f^*m}} \left(e^{-(\Phi_m - \Phi_{f^*m})z_2^{-\theta}} e^{-\Phi_{f^*m} z_1^{-\theta}} - e^{-\Phi_m z_2^{-\theta}}\right).
\end{aligned}$$

The second part of equation (A-2) is therefore

$$\Pr\left(z_2 \le \tilde{Z}_{1m}(i) \le z_1, \tilde{Z}_{2m}(i) \le z_2 \mid \tilde{Z}_{1m}(i) > V\right) = \frac{\Phi_m}{\Phi_{f^*m}} \left(e^{-(\Phi_m - \Phi_{f^*m})z_2^{-\theta}} e^{-\Phi_{f^*m} z_1^{-\theta}} - e^{-\Phi_m z_2^{-\theta}}\right). \tag{A-4}$$

Summing equation (A-3) and (A-4), the joint distribution of highest two cost-adjusted productivities conditional on f^* from ℓ^* selling to i in m is

$$F_{12,m}(z_1, z_2; \ell^*, f^*) = \frac{\Phi_m}{\Phi_{f^*m}} e^{-(\Phi_m - \Phi_{f^*m})z_2^{-\theta}} e^{-\Phi_{f^*m} z_1^{-\theta}} - \frac{\Phi_m - \Phi_{f^*m}}{\Phi_{f^*m}} e^{-\Phi_m z_2^{-\theta}}.$$

The associated p.d.f. is

$$f_{12,m}(z_1, z_2; \ell^*, f^*) = \Phi_m(\Phi_m - \Phi_{f^*m})\theta^2 z_1^{-\theta-1} z_2^{-\theta-1} e^{-(\Phi_m - \Phi_{f^*m})z_2^{-\theta}} e^{-\Phi_{f^*m} z_1^{-\theta}}.$$

A.2 Price distribution

The price of consumer i in market m is

$$P_m(i) = \min\{\frac{1}{\tilde{Z}_{2m}(i)}, \frac{\bar{\mu}}{\tilde{Z}_{1m}(i)}\}.$$

Conditional on firm f^* serves the consumer in the market, the complement of the price c.d.f. is

$$1 - F_m^P(p; f^*) = \underbrace{\Pr\left(p \leq \frac{1}{\tilde{Z}_{2m}(i)} < \frac{\bar{\mu}}{\tilde{Z}_{1m}(i)} \mid \tilde{Z}_{1m}(i) > V\right)}_{\text{T1}} + \underbrace{\Pr\left(p \leq \frac{\bar{\mu}}{\tilde{Z}_{1m}(i)} \leq \frac{1}{\tilde{Z}_{2m}(i)} \mid \tilde{Z}_{1m}(i) > V\right)}_{\text{T2}}.$$

Derive each component, I have the firm term

$$\begin{aligned} \text{T1} &= \int_{p^{-1}}^{\infty} \int_{z_1/\bar{\mu}}^{p^{-1}} f_{12,m} dz_2 dz_1 + \int_{0}^{p^{-1}} \int_{z_1/\bar{\mu}}^{z_1} f_{12,m} dz_2 dz_1 \\ &= \frac{\Phi_m}{\Phi_{f^*m}} e^{-(\Phi_m - \Phi_{f^*m})p^\theta} - \frac{\Phi_m - \Phi_{f^*m}}{\Phi_{f^*m}} e^{-\Phi_m p^\theta} - \frac{\Phi_m}{\Phi_{f^*m} + (\Phi_m - \Phi_{f^*m})\bar{\mu}^\theta}, \end{aligned}$$

and the second term

$$\begin{aligned} \text{T2} &= \int_{0}^{\infty} \int_{\bar{\mu} z_2}^{\bar{\mu}/p} f_{12,m} dz_1 dz_2 \\ &= \frac{\Phi_m}{\Phi_{f^*m}} e^{-\Phi_{f^*m}\bar{\mu}^{-\theta} p^\theta} - \frac{\Phi_m/\Phi_{f^*m}(\Phi_m - \Phi_{f^*m})\bar{\mu}^\theta}{\Phi_{f^*m} + (\Phi_m - \Phi_{f^*m})\bar{\mu}^\theta}. \end{aligned}$$

Combining the two and subtracted by one, I get the price distribution exactly equals to equation (6).

A.3 Markup distribution

The markup equals

$$\mu_m(i) = \min\left\{\bar{\mu}, \frac{C_{2m}(i)}{C_{1m}(i)}\right\}.$$

Conditional on firm f^* serves the consumer i in market m, for the range below the monopoly markup, the distribution is

$$F_m^\mu(\mu; f^*) = \Pr\left(\frac{\tilde{Z}_{1m}(i)}{\tilde{Z}_{2m}(i)} \leq \mu \mid \tilde{Z}_{1m}(i) > V\right).$$

Let's first calculate the complement of the c.d.f.,

$$
\begin{aligned}
1 - F_m^\mu(\mu; f^*) &= \Pr\left(\tilde{Z}_{2m}(i) \leq \mu^{-1}\tilde{Z}_{1m}(i) \mid \tilde{Z}_{1m}(i) > V\right) \\
&= \int_0^\infty \int_0^{\mu^{-1}z_1} f_{12,m}(z_1, z_2; f^*) dz_2 dz_1 \\
&= \frac{\Phi_m}{\Phi_{f^*m} + (\Phi_m - \Phi_{f^*m})\mu^\theta}.
\end{aligned}
$$

The conditional markup distribution is then

$$
F_m^\mu(\mu; f^*) = 1 - \frac{1}{\mu^\theta - \frac{\Phi_{f^*m}}{\Phi_m}(\mu^\theta - 1)} = 1 - \frac{1}{(1 - s_{f^*m})\mu^\theta + s_{f^*m}},
$$

where $s_{f^*m} = \Phi_{f^*m}/\Phi_m$. Given the markup $\mu \in (1, \infty)$, it's obvious that $\lim_{\mu\to1} F_m^\mu(\mu; f^*) = 0$ and $\lim_{\mu\to\infty} F_m^\mu(\mu; f^*) = 1$.

The markup distribution is truncated at the monopoly markup,

$$
F_m^\mu(\mu; f^*) = \begin{cases} 1 - \frac{1}{(1-s_{f^*m})\mu^\theta + s_{f^*m}} & 1 \leq \mu < \bar{\mu} \\ 1 & \mu \geq \bar{\mu} \end{cases}.
$$

The markup increases with the number of locations a firm builds plants. Moreover, I will show below that the probability of a firm earning monopoly markup increases with its number of plants.

Define $F_m^\mu(\mu; f^*, z_2)$ as the probability that $1 \leq \frac{\tilde{Z}_{1m}(i)}{\tilde{Z}_{2m}(i)} \leq \mu$, given the second-lowest cost and firm f^* wins the consumer. It can be simplified as

$$
\begin{aligned}
F_m^\mu(\mu; f^*, z_2) &= \Pr\left(\tilde{Z}_{2m}(i) \leq \tilde{Z}_{1m}(i) \leq \mu\tilde{Z}_{2m}(i) \mid \tilde{Z}_{2m}(i) = z_2\right) \\
&= \frac{\int_{z_2}^{\mu z_2} f_{12,m}(z_1, z_2) dz_1}{\int_{z_2}^{\infty} f_{12,m}(z_1, z_2) dz_1} \\
&= \frac{e^{-\Phi_{f^*m}(\mu z_2)^{-\theta}} - e^{-\Phi_{f^*m} z_2^{-\theta}}}{1 - e^{-\Phi_{f^*m} z_2^{-\theta}}}.
\end{aligned}
$$

Therefore, the probability of firm f^* charging monopoly markup is

$$
1 - F_m^\mu(\bar{\mu}; f^*, z_2) = \frac{1 - e^{-\Phi_{f^*m}(\bar{\mu} z_2)^{-\theta}}}{1 - e^{-\Phi_{f^*m} z_2^{-\theta}}}.
$$

Taking first order derivative with respect to Φ_{f^*m}, we see that the probability of the firm earning monopoly markup strictly increases with its producing capability Φ_{f^*m}. Since Φ_{f^*m} increases with firm's number of plants, it implies that the more plants a firm builds, the higher likely it can charge monopoly markup.

A.4 Expected revenue

Before calculating the expected revenue and cost, it is useful to state the Gamma Lemma proved in appendix 5.1 of Holmes et al. (2011).

Gamma Lemma:

(i) For $\omega > 0$ and $\theta - \eta + 1 > 0$,

$$\int_0^\infty z^{\eta-\theta-2} e^{-\omega z^{-\theta}} dz = \omega^{\frac{\eta-\theta-1}{\theta}} \frac{1}{\theta} \Gamma\left(\frac{\theta-\eta+1}{\theta}\right)$$

(ii) For $\omega > 0$ and $2\theta - \eta + 1 > 0$,

$$\int_0^\infty z^{\eta-2\theta-2} e^{-\omega z^{-\theta}} dz = \omega^{\frac{\eta-2\theta-1}{\theta}} \left(\frac{\theta-\eta+1}{\theta^2}\right) \Gamma\left(\frac{\theta-\eta+1}{\theta}\right).$$

The conditional expected revenue is

$$E\big[R_{fm} \mid f = f^*\big] = A_m E\big[p_m(i)^{1-\eta}\big],$$

which is the expected revenue for cement sold to destination market m, conditional on sourcing from firm f^*, and fixing firm f^*'s plant locations. The expectation is taken with respect to the random price realization. The demand shifter $A_m = \exp(\alpha_m)$.

For $p_m(i) = \min\left\{\frac{1}{\tilde{Z}_{2m}(i)}, \frac{\bar{\mu}}{\tilde{Z}_{1m}(i)}\right\}$, we have the expectation

$$E\big[p_m(i)^{1-\eta}\big] = \underbrace{\int_0^\infty \int_{\frac{z_1}{\bar{\mu}}}^{z_1} \left(\frac{1}{z_2}\right)^{1-\eta} f_{12,m}(z_1,z_2) dz_2 dz_1}_{\text{T1}} + \underbrace{\int_0^\infty \int_0^{\frac{z_1}{\bar{\mu}}} \left(\frac{\bar{\mu}}{z_1}\right)^{1-\eta} f_{12,m}(z_1,z_2) dz_2 dz_1}_{\text{T2}}.$$

The first term can be simplified after changing the order of integration and applying the Gamma Lemma,

$$\text{T1} = \frac{\Phi_m}{\Phi_{f^*m}} \left(\Phi_m - \Phi_{f^*m}\right) \Gamma\left(\frac{\theta-\eta+1}{\theta}\right) \left[\left(\Phi_m - (1-\bar{\mu}^{-\theta})\Phi_{f^*m}\right)^{\frac{\eta-\theta-1}{\theta}} - \Phi_m^{\frac{\eta-\theta-1}{\theta}}\right].$$

The second term can be simplified to

$$\text{T2} = \bar{\mu}^{-\theta} \Phi_m \Gamma\left(\frac{\theta-\eta+1}{\theta}\right) \left(\Phi_m - (1-\bar{\mu}^{-\theta})\Phi_{f^*m}\right)^{\frac{\eta-\theta-1}{\theta}}.$$

Combining the two terms, I have the conditional expected revenue equals to

$$E\big[R_{fm} \mid f = f^*\big] = A_m \Gamma\left(\frac{\theta - \eta + 1}{\theta}\right) \frac{\hat{R}_{f^*m}}{s_{f^*m}},$$

where $\hat{R}_{f^*m} = \left(\Phi_m - (1 - \bar{\mu}^{-\theta})\Phi_{f^*m}\right)^{\frac{\eta-1}{\theta}} - \left(\Phi_m - \Phi_{f^*m}\right)\Phi_m^{\frac{\eta-\theta-1}{\theta}}$, and $s_{f^*m} = \Phi_{f^*m}/\Phi_m$. The unconditional expected revenue is therefore,

$$E\big[R_{fm}\big] = A_m \kappa \hat{R}_{fm}, \text{ where } \kappa = \Gamma\left(\frac{\theta - \eta + 1}{\theta}\right).$$

A.5 Costs

The conditional expected cost of a firm is

$$E\big[C_{fm} \mid f = f^*\big] = A_m E\left[\frac{p_m(i)^{1-\eta}}{\mu}\right],$$

for $\mu = \frac{\tilde{Z}_{1m}(i)}{\tilde{Z}_{2m}(i)}$ when $p_m(i) = \frac{1}{\tilde{Z}_{2m}(i)}$ and $\mu = \bar{\mu}$ when $p_m(i) = \frac{\bar{\mu}}{\tilde{Z}_{1m}(i)}$. Hence,

$$E\left[\frac{p_m(i)^{1-\eta}}{\mu}\right] = \underbrace{\int_0^\infty \int_{\frac{z_1}{\bar{\mu}}}^{z_1} \left(\frac{1}{z_2}\right)^{1-\eta} \frac{z_2}{z_1} f_{12,m}(z_1, z_2) dz_2 dz_1}_{\text{T1}} + \underbrace{\int_0^\infty \int_0^{\frac{z_1}{\bar{\mu}}} \left(\frac{\bar{\mu}}{z_1}\right)^{1-\eta} \frac{1}{\bar{\mu}} f_{12,m}(z_1, z_2) dz_2 dz_1}_{\text{T2}}.$$

The simplification of the firm term is more involved. I need to replace z_1 by μz_2 and change the order of integration (refer to the appendix of Holmes et al. (2011)). The first term equals to

$$\text{T1} = \Phi_m(\Phi_m - \Phi_{f^*m})(\theta - \eta + 1)\Gamma\left(\frac{\theta - \eta + 1}{\theta}\right) \int_1^{\bar{\mu}} \mu^{-\theta-2} \left(\Phi_m - (1 - \mu^{-\theta})\Phi_{f^*m}\right)^{\frac{\eta-2\theta-1}{\theta}} d\mu.$$

Unfortunately, there is no closed-form expression for the integral. Therefore, I apply the numerical approximation in the empirical section.

Applying the Gamma Lemma and taking the same steps as deriving the second term in the expected revenue function, the second term here can be simplified to

$$\text{T2} = \bar{\mu}^{-\theta-1} \Phi_m \Gamma\left(\frac{\theta - \eta + 1}{\theta}\right) \left(\Phi_m - (1 - \bar{\mu}^{-\theta})\Phi_{f^*m}\right)^{\frac{\eta-\theta-1}{\theta}}.$$

Combining the two terms, I derive the conditional expected cost equals to

$$E\big[C_{fm} \mid f = f^*\big] = A_m \Gamma\left(\frac{\theta - \eta + 1}{\theta}\right) \frac{\hat{C}_{f^*m}}{s_{f^*m}},$$

where

$$\hat{C}_{f^*m} = \Phi_{f^*m}\Bigg[(\theta-\eta+1)(\Phi_m-\Phi_{f^*m})\int_1^{\bar{\mu}} \mu^{-\theta-2}\left(\Phi_m-(1-\mu^{-\theta})\Phi_{f^*m}\right)^{\frac{\eta-2\theta-1}{\theta}} d\mu \\ + \bar{\mu}^{-\theta-1}\left(\Phi_m-(1-\bar{\mu}^{-\theta})\Phi_{f^*m}\right)^{\frac{\eta-\theta-1}{\theta}}\Bigg].$$

The unconditional expected cost is therefore

$$E\big[C_{fm}\big] = A_m\kappa\hat{C}_{fm}.$$

A.6 Best-response potential game

A best-response potential game is where potential functions infer the difference in the payoff due to unilaterally deviation of each player to the best response. It is introduced in work of Monderer and Shapley (1996) and later on developed in Voorneveld (2000). Under the condition of a finite game where the number of players is finite and each of them has a finite strategy space, a best-response potential game always has pure strategy Nash equilibrium and more interestingly, every learning process based on best-response of the players converges to an Nash equilibrium.

Moreover, starting from any arbitrary location decisions, if players simultaneously deviates to their unique best replies in each period, the process terminates in a Nash equilibrium after finite number of steps. Swenson and Kar (2017) found that the convergence rate is exponential. Table A.4 shows examples of 6 to 12 locations and two firms. Each is solved for 1000 times. The maximum number of rounds to find an equilibrium is three. When the potential number of locations is larger and therefore the strategy space is larger, it takes longer to find an equilibrium but still converges to a solution relatively fast.

Table A.4: Convergence Rate Check of Best-Response Potential Game

Number of locations	Average time (seconds)	Average number of BR rounds	Max number of BR rounds
6	0.0198	1.0830	3
7	0.0429	1.1010	2
8	0.0494	1.0190	2
9	0.0596	1.1830	3
10	0.0934	1.1230	3
11	0.0963	1.1980	2
12	0.1275	1.1130	2

B Model extensions

B.1 Adding core productivity differences at firm level

Suppose the firm's endowed core productivity also characterizes its plants' marginal cost of production,

$$C_{f\ell m}(i) = \frac{w_\ell \tau_{\ell m}}{Z_f Z_\ell(i)},$$

where $Z_\ell(i)$ are draws from the Fréchet distribution $\exp(-T_\ell z^{-\theta})$, and Z_f are firm-specific parameters.

The c.d.f. of the plant's cost-adjusted productivity $\tilde{Z}_{f\ell m}(i) = \frac{Z_\ell(i)}{w_\ell \tau_{\ell m}/Z_f}$ is then

$$\tilde{F}^{draw}_{f\ell m}(z) = \exp(-\phi_{f\ell m} z^{-\theta}),$$

where $\phi_{f\ell m} = Z_f^\theta \phi_{\ell m} = Z_f^\theta T_\ell \left(w_\ell \tau_{\ell m}\right)^{-\theta}$. The distributions of plants' productivities at the same location are shifted by firms' core productivities, although the shape parameter remains the same. Plants owned by an efficient firm are on average more productive than those owned by inefficient firms at the same location. Exploiting the properties of extreme value distribution, the distribution of a firm's highest cost-adjusted productivity in supplying the product to market m is

$$\tilde{F}_{1,fm}(z) = \exp(-\Phi_{fm} z^{-\theta}),$$

where $\Phi_{fm} = \sum_\ell \mathbb{I}_{f\ell} \phi_{f\ell m}$. The firm's capability not only depends on plants spatial setting but also its core productivity.

Other than the difference in the formulation of Φ_{fm}, what followed in completing the multi-plant firm model all remains the same. Specifically, the probability that a location exports good to a market becomes

$$s_{\ell m} = \frac{\sum_f \mathbb{I}_{f\ell} \phi_{f\ell m}}{\Phi_m}.$$

Transforming the sourcing probabilities into the gravity-type regression, one gets the same form as in equation (14), but with the location fixed effects being $\text{FE}_\ell = \ln\left(T_\ell w_\ell^{-\theta} \sum_f \mathbb{I}_{f\ell} Z_f^\theta\right)$. Therefore, one can no longer separately identify the location characteristics $T_\ell w_\ell^{-\theta}$ from the firm productivities Z_f without the help of additional firm-level data.

The gravity model, however, still holds at plant level where $s_{f\ell m} = \frac{\phi_{f\ell m}}{\Phi_m}$ conditional on firm f has a plant at location ℓ, and the estimable form is

$$E\left[\frac{Q_{f\ell m}}{Q_m} \mid \mathbb{I}_{f\ell} = 1\right] = \exp\left[\text{FE}_f + \text{FE}_\ell + \text{FE}_m - \theta \mathbf{X}'_{\ell m} \beta^\tau\right],$$

where $\text{FE}_f = \theta \ln Z_f$ and $\text{FE}_\ell = \ln\left(T_\ell w_\ell^{-\theta}\right)$. Plant-market-level trade flow in volume will be needed in performing the first step of the estimation.

C Simulation details

C.1 Delta Method for transformed parameters in SP approximation

From equation (20), we have the transformed parameters of interest as

$$G(\mathbf{b}) = \begin{cases} \widehat{\mu_B^F} & = -\frac{b(\mathbb{I}\{f=B\})}{b(\ln E[\pi_{f\ell}])} \\ \widehat{\mu_M^F} & = -\frac{b(\mathbb{I}\{f=M\})}{b(\ln E[\pi_{f\ell}])} \\ \widehat{\sigma^F} & = \frac{1}{b(\ln E[\pi_{f\ell}])}, \end{cases}$$

where $b(\ln E[\pi_{f\ell}])$ is the estimated coefficient of $\ln E[\pi_{f\ell}]$, and $b(\mathbb{I}\{f = \cdot\})$ are coefficients of respective firm dummy.

Using delta method, the approximate variance of the transformed $\left(\widehat{\mu_B^F}, \widehat{\mu_M^F}, \widehat{\sigma^F}\right)$ is

$$V(G(\mathbf{b})) = J(G(\mathbf{b}))V(\mathbf{b})J(G(\mathbf{b}))',$$

where $\mathbf{b}$ is the vector of coefficients and $V(\mathbf{b})$ is their variance covariance matrix. $J(G(\mathbf{b}))$ is the Jacobian matrix of $G(\mathbf{b})$ with respect to $\mathbf{b}$, which is

$$J(G(\mathbf{b})) = \begin{bmatrix} -\frac{1}{b(\ln E[\pi_{f\ell}])} & 0 & \frac{b(\mathbb{I}\{f=B\})}{b(\ln E[\pi_{f\ell}])^2} \\ 0 & -\frac{1}{b(\ln E[\pi_{f\ell}])} & \frac{b(\mathbb{I}\{f=M\})}{b(\ln E[\pi_{f\ell}])^2} \\ 0 & 0 & -\frac{1}{b(\ln E[\pi_{f\ell}])^2}. \end{bmatrix}$$