Plants in Space*

Ezra Oberfield
Princeton University

Esteban Rossi-Hansberg
Princeton University

Pierre-Daniel Sarte
Federal Reserve Bank of Richmond

Nicholas Trachter
Federal Reserve Bank of Richmond

June 30, 2020

Abstract

We study the number, size, and location of a firm's plants. The firm's decision balances the benefit of delivering goods and services to customers using multiple plants with the cost of setting up and managing these plants, and the potential for cannibalization that arises as their number increases. Modeling the decisions of heterogeneous firms in an economy with a vast number of widely distinct locations is complex because it involves a large combinatorial problem. Using insights from discrete geometry, we study a tractable limit case of this problem in which these forces operate at a local level. Our analysis delivers clear predictions on sorting across space. Productive firms place more plants in dense locations that exhibit high rents compared with less productive firms, and place fewer plants in markets with low density and low rents. Controlling for the number of plants, productive firms also operate larger plants than those operated by less productive firms in locations where both are present. We present evidence consistent with these and several other predictions using U.S. establishment-level panel data.

*Oberfield: edo@princeton.edu. Rossi-Hansberg: erossi@princeton.edu. Sarte: pdgs4frbr@gmail.com. Trachter: trachter@gmail.com. We thank participants at numerous seminars and conferences for their feedback. We thank Eric LaRose, Reiko Laski, James Lee, Sara Ho, and Suren Tavakalov for outstanding research assistance. The views expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1 Introduction

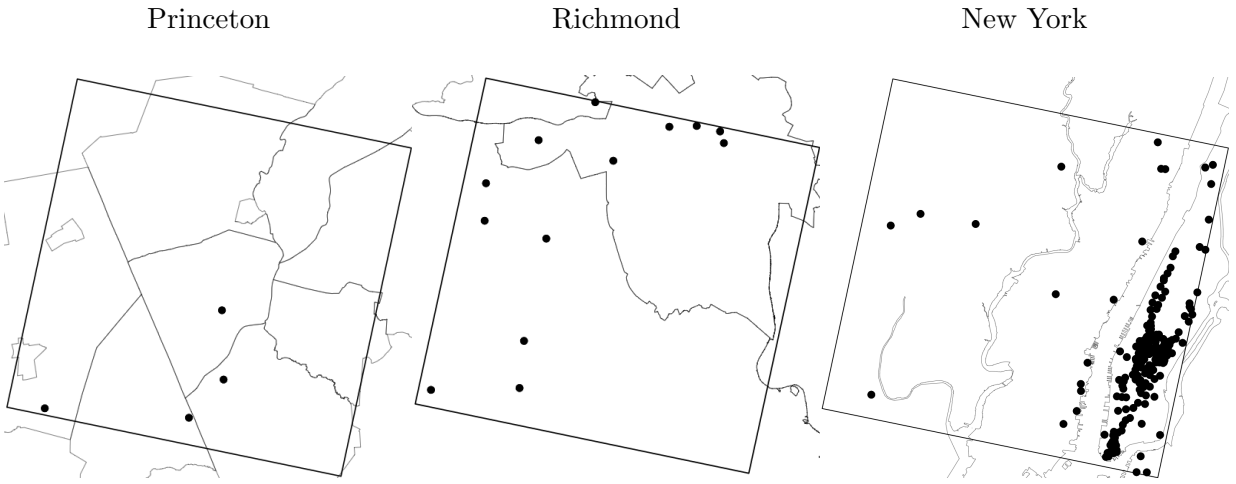
Delivering products and services to locations where consumers can easily access them involves complex decisions on where to locate plants and how large these plants should be. Having too few establishments in the wrong locations is costly because it increases either transport costs or the distance to consumers. Having too many involves large span-of-control and fixed costs, as well as plants that cannibalize each other's customers. Understanding how these trade-offs play out for firms with different characteristics in an economy consisting of many local markets that differ in demand and production costs is complex. Perhaps due to the difficulty of the problem, little is known about the solution to this fundamental problem of how to organize production. The sorting of firms in space determines not only the profitability of firms, but also consumers' surplus as well as the characteristics of individual locations. In this paper, we study this core component of a firm's production problem, provide a methodology that simplifies it significantly, and contrast its implications with the data.

Consider the case of Starbucks, which operated in 2019 around 14,000 stores in different locations across the US. Of course, not all Starbucks are equal in size, not all locations in the US have a Starbucks, and the distance between neighboring Starbucks stores in a location differs across space. Simply put, there is a lot of variation across space in how individual stores are arranged. This variation is naturally related to the spatial distribution of population density, of wages, and other characteristics. For example, Figure 1 shows the location of Starbucks establishments in three markets, Princeton NJ, Richmond VA, and New York NY. Clearly, the number of establishments as well as the distance between them varies across these cities. Even within New York, the number of establishments is much larger, and the distance between them much shorter, in the densest parts of Manhattan. What are the general characteristics of establishment location decisions? Clearly density matters, but the scale of establishments is by no means constant in space. The average plant employment of Starbucks in New York is more than 23% higher than in Richmond.

Casual evidence and introspection might suggest that firms simply sell in the densest markets with the marginal market determined by a firm's productivity. A closer look, however, reveals a more nuanced pattern. Figure 2 provides a simple example that illustrates this point. Walgreens and Rite Aid are pharmacies that operate nationally, but Walgreens' total employment is larger and it has more establishments. The figure shows that, in fact, both pharmacy chains tend to have more establishments in more dense locations. However, Rite Aid decided to have more stores in less dense locations. Is this form of sorting across location a general feature of the solution to the location problem, of the data?

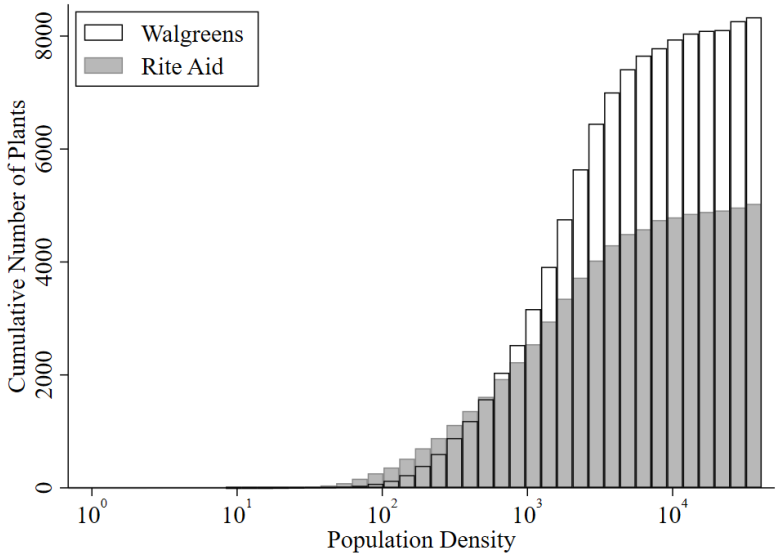
More generally, we aim to provide insights into two main questions: First, which firms set up plants in which locations? Second, what determine the scale and location of production? Answering these questions requires us to think about plants and firms as distinct, albeit related, economic entities. In particular, we set up an economy with a continuum of heterogeneous locations. These locations have different productivities and amenities that determine, in general equilibrium, the distribution of population density, wages, and residential and business rents. It also determines, given the form local competition takes in a location, residual

Figure 1: Density of Starbucks Locations



Notes: The figure shows the locations of Starbucks establishments in a 12×12 square mile area in Princeton, Richmond, and New York City.

Figure 2: Sorting: Walgreens vs. Rite Aid



Notes: For Walgreens and Rite Aid, this graph plots the cumulative number of establishments in locations with populations density weakly less than each population density. Population density is measured as the population density of the 6×6 mile square in which the establishment is located.

demand for a firm's product. We focus primarily on the problem of a firm that takes these distributions as given and needs to decide if and how to serve each consumer. The firm decides where to set up production plants, how large each plant should be, and from which plant to serve each of its customers. We assume that firms face iceberg transport costs. Setting up a plant entails a fixed cost that depends on local land rents. The productivity of a plant depends on local characteristics as well as a firm-specific component that decreases with the total number of plants the firm operates. In other words, increasing a firm's span-of-control by adding plants implies a management cost that lowers its productivity. The main trade-off faced by the firm, therefore, is to reduce transport costs by setting up more plants close to consumers versus setting up fewer plants to economize on fixed costs, augment productivity by lowering span of control, and reduce cannibalization between plants. Ours is a standard setup of this canonical firm decision problem.

Solving this production problem when the set of potential locations is large and heterogeneous involves a large and challenging combinatorial optimization problem. Our contribution is to focus on a limit formulation of the problem in which the firm chooses a density of plants, rather than a discrete set. The firm's problem then becomes one of calculus of variations which is simpler to solve. Crucially, in the limit we propose, all relevant trade-offs described above remain active. Specifically, we study a limit of the problem in which fixed and span-of-control managerial costs become small while transportation costs become large. In this limit economy, the problem of the firm becomes amenable to an analytical characterization, making it easy to transparently characterize its implications.

In characterizing the solution to the firm's problem in the limit economy, we apply insights from discrete geometry. In particular, we exploit the sum of moments theorem by Fejes Toth (1953). The theorem provides the optimal way to arrange plants across space when economic activity is uniform and the number of plants is large. In particular, it states that plants should be located at the center of catchment areas given by hexagons arranged so as to cover all locations. The intuition for this result is that, among all polygons with which one can construct a uniform tessellation, the hexagon is the closest to a circle.[1] A circle minimizes the average distance from a plant located at is center to its customers. However, unlike hexagons, it cannot be used repeatedly to form a tessallation. We extend the theorem to an economy where locations are heterogeneous across space. Specifically, the environment recognizes that customers are not necessarily uniformly distributed across space while plant costs and productivities also differ across locations.

Apart from being obviously important for practical applications, our extension of the theorem allows us to study sorting patterns, namely the many-to-many matching between heterogeneous firms and heterogeneous locations. It helps us understand examples such as Starbucks, or Walgreens and Rite Aid, for which the number of establishments changes with customer density but at different rates and with different ranges of locations that vary with the firm's aggregate scale. In general, our theory delivers testable implications on sorting patterns across firms. First, more productive firms set up *more* plants in denser, high rent, locations than low productivity firms. Perhaps more surprising is that they also set up *fewer* plants in markets with

[1] A tessellation is an arrangement of shapes, especially of polygons, in a repeated pattern without gaps or overlapping.

lower density and rents. Second, controlling for the number of plants in a location, more productive firms set up larger plants. In the final section of the paper, we present evidence, using the National Establishment Time Series (NETS) dataset, that corroborates these and other predictions of our theoretical analysis.

One of the advantages of using our limit economy to analyze the firm production problem is that the simplicity of the solution allows us to embed the problem into an equilibrium setup. In equilibrium, exogenous local amenities and populations, together with worker mobility and the solution to the firm's production problem, lead to an endogenous distribution of population in space, as well as a distribution of local prices and factor costs. We set up this problem and show that in equilibrium, an index of local productivity and amenities is the only relevant local characteristic needed to determine equilibrium outcomes. Analyzing realistic quantitative general equilibrium counterfactuals is not the focus of our study, but the method we develop to make the firm's problem tractable can be readily used to do so. We provide an algorithm to compute an industry equilibrium of our economy and illustrate the effect of improvements, in a single industry, in the technology to manage a firm's span of control and the technology to transport goods.

Canonical models of firm dynamics (i.e. Jovanovic (1982) and Hopenhayn (1992)) make no clear distinction between a plant and a firm. However, mounting evidence points to the importance of considering plants and firms as different but related entities. Rossi-Hansberg and Wright (2007) highlight large differences between the size distributions of enterprises and establishments. In addition, Rossi-Hansberg et al. (2018) show evidence of diverging trends in market concentration at the national and local levels resulting from the expansion of the largest firms into new markets. Hsieh and Rossi-Hansberg (2019) show that industries with large increases in national market concentration also saw their top firms expand their operations geographically through the opening of new plants in smaller markets. Further, Aghion et al. (2019) observe that the average number of plants per firm has risen considerably in the US, and Cao et al. (2019) and Aghion et al. (2019) provide evidence that growth through the opening of new plants has been a key margin of a firm's employment growth since 1990.

The distinction between firms and plants has been more prevalent in the international trade literature given the interest in multinational production and export platforms. Examples of papers in this literature include Ramondo (2014), Ramondo and Rodríguez-Clare (2013), and Tintelnot (2016), among many others. Moreover, all of these frameworks either solve the combinatorial problem with only a few countries or assume away fixed costs. These models all have the feature that the less profitable markets are reached only by the more productive firms. In contrast, our environment is one in which it is the less productive firms that locate in the more marginal markets.

The industrial organization literature has also analyzed how individual firms set up distribution networks in space. Seminal papers include Jia (2008) and Holmes (2011). Importantly, many of these frameworks study cases where opening stores in one location increases the marginal value of opening stores in other locations, the so called "supermodular" cases.[2] The lack of cannibalization across plants makes these cases somewhat easier to handle. On the contrary, cases where cannibalization is prevalent, so that setting up

[2]Holmes (2011) assumes submodularity but does not solve the model; he estimates parameters using moment inequalities.

new plants reduces the value of other plants, cannot easily be solved except for algorithms that evaluate essentially all possible combinations. Recently, Arkolakis and Eckert (2017) and Hu and Shi (2019) have developed more powerful algorithms to solve these types of "submodular" problems more efficiently, but doing so for large numbers of locations remains a challenge. Furthermore, the purely numerical nature of essentially all this literature implies that few general insights have been obtained. Our analytical approach has the advantage of providing a set of general implications that we can contrast with micro data.

We provide a characterization of the matching of heterogeneous firms with multiple plants to heterogeneous locations. Gaubert (2018), and Ziv (2019) study the assignment of single-plant firms to heterogeneous locations. Behrens et al. (2014), Eeckhout et al. (2014), Diamond (2016), Davis and Dingel (2019) and Bilal and Rossi-Hansberg (2019), study the assignment of workers to heterogeneous locations. None of these papers, however, address sorting when the agent, in our case a firm, can choose many locations concurrently.[3]

The rest of the paper is organized as follows. Section 2 presents the problem of the firm, proposes and studies the limit problem, and derives our main results. Section 3 embeds heterogeneous firms solving the production problem with multiple plants into a general equilibrium economy. It also presents numerical examples that illustrate the effect of changes in the efficiency of span of control and transport costs. Section 4 contrasts some of the main implications of our solution with panel data of firms and establishments. Section 5 concludes. An Appendix includes technical derivations, presents additional robustness results and data constructions details, and describes the numerical algorithm.

2 The Multi-plant Firm Problem

We consider the problem of a firm deciding how to serve customers located in a unit square, $\mathcal{S} = [0,1]^2 \subset \mathbb{R}^2$. More generally, we require $\mathcal{S}$ to be an Euclidean space that is closed, bounded and convex. Each location $s \in \mathcal{S}$ is characterized by an exogenous productivity level B_s, as well as local equilibrium characteristics that firms take as given, namely, the residual demand function, $D_s(\cdot)$, the wage rate, W_s, and the commercial rent, R_s.

There is a set of firms, $j \in J$. Each firm produces a unique variety. A firm is characterized by productivity, q_j. It chooses a finite set of locations $O_j \subset \mathcal{S}$ where to set up plants. If j produces in a total of $N_j = |O_j|$ locations, its productivity in location $o \in O_j$ is $B_o Z(q_j, N_j)$ where Z is decreasing in N and $Z(q,0) < \infty$. Thus, firms face a cost of increasing their span of control. Each plant takes up ξ units of commercial real estate with rental cost R_s per unit of space. Trade between any two locations incurs an iceberg shipping cost. For one unit of a good to arrive a distance of δ, $T(\delta) \geq 1$ units must be shipped. We assume $T(\delta)$ is strictly increasing, satisfies the triangle inequality, and diverges as $\delta \to \infty$.

Conditional on having a plant, production requires only local labor which is employed at wage W_s. Let

[3] Empirically, assessing sorting patterns when each plant is a stand alone unit is difficult because of the reflection problem. In particular, one can observe whether plants in more dense locations are larger, but it is not clear whether that is due to sorting or to the impact of being in a dense location. In our setting, with firms that operate many units in different locations, we can exploit leave-out strategies to argue that there is clear evidence of positive assortative matching.

l_{jo} denote the number of workers employed by firm j at a plant in o. Firms will serve customers in the least costly possible way. Thus, the cost of delivering one unit of good j from a plant in o to a consumer in s is $\frac{W_o T(\delta_{so})}{B_o Z(q_j,N_j)}$. Let $\Lambda_{js}(O_j) \equiv \min_{o\in O_j} \frac{W_o T(\delta_{so})}{B_o Z(q_j,N_j)}$ be j's minimal cost of delivering one unit of good j to a consumer in location s, and let $\sigma_{js} \equiv \arg\min_{o\in O_j} \frac{W_o T(\delta_{so})}{B_o Z(q_j,N_j)}$ be the location that sources the product. Let p_{js} be the price charged by j to consumers in s. Then, $D_s(p_{js})$ be the residual demand for variety j in location s. The optimal price maximizes

$$\max_{p_{js}} D_s(p_{js})\left(p_{js} - \Lambda_{js}\right) \ ,$$

The problem above can lead to complicated pricing rules where markups depend on local characteristics. To simplify the problem, we abstract from spatial variation in markups and assume the following about the residual demand function.

Assumption 1 *Residual demand satisfies $D_s(p_{js}) = D_s p_{js}^{-\varepsilon}$, where D_s subsumes all determinants of local demand, including the local price index.*

Assumption 1 is satisfied in the standard case with monopolistic competition and CES preferences with elasticity of substitution across varieties given by ε. Then, as usual, $p_{js} = \frac{\varepsilon}{\varepsilon-1}\Lambda_{js}$.[4]

Firm j's profit can be expressed as

$$\pi_j = \max_{O_j}\left\{\int_s \max_{p_{js}} D_s p_{js}^{-\varepsilon}\left(p_{js} - \Lambda_{js}\left(O_j\right)\right) ds - \sum_{o\in O_j} R_o\xi\right\}, \tag{1}$$

or, using the expression for j's price, is

$$\begin{aligned}
\pi_j &= \max_{O_j}\left\{\frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}}\int_s D_s \Lambda_{js}\left(O_j\right)^{1-\varepsilon} ds - \sum_{o\in O_j} R_o\xi\right\} \\
&= \max_{O_j}\left\{\frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}}\int_s D_s\left(\min_{o\in O_j}\left\{\frac{W_o T\left(\delta_{so}\right)}{B_o Z\left(q_j,N_j\right)}\right\}\right)^{1-\varepsilon} ds - \sum_{o\in O_j} R_o\xi\right\} \\
&= \max_{O_j}\left\{Z\left(q_j,N_j\right)^{\varepsilon-1}\frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}}\int_s D_s \max_{o\in O_j}\left\{\left(\frac{B_o/W_o}{T\left(\delta_{so}\right)}\right)^{\varepsilon-1}\right\} ds - \sum_{o\in O_j} R_o\xi\right\}.
\end{aligned}$$

[4] In this environment, the fact that a plant can charge different prices to different customers is innocuous because markups are independent of demand. This would be equivalent to having the store charging a single price to all its customers, but customers incurring on the iceberg cost of transporting the good back to their home.

2.1 The Catchment Area of a Plant

The catchment area of a plant in location o is formed by locations s to which the firm ships goods from the plant in o. Formally, letting $CT(o)$ denote the catchment area of a plant in location o,

$$CT(o) = \left\{ s \in \mathcal{S} \text{ for which } o = \arg\max_{\tilde{o} \in O_j} \left\{ \frac{B_{\tilde{o}}/W_{\tilde{o}}}{T\left(\delta_{s\tilde{o}}\right)} \right\} \right\} . \tag{2}$$

Given that $T(0) = 1$ and T is increasing, if $CT(o)$ is not empty then $o \in CT(o)$. $CT(o)$ can be empty if a plant's cost, relative to other nearby plants, is high enough.[5]

Notice that, once plants are placed in locations O_j, the catchment area of each plant only depends on transportation costs and on the production cost of locations where the plants are placed. The production cost plays an important role in defining catchment areas.

2.1.1 An example

We explore the importance of the production cost by solving for the catchment areas when we place 9 plants equidistant from each other.[6] For transportation costs we let $T(\delta_{so}) = 1 + \delta_{so}$, where δ_{so} is the Euclidean distance between s and o. We solve two cases with different distributions of productivity and economic activity. In the first case, we assume that there is no variation in economic fundamentals, so $B_o/W_o = 1$ $\forall\ o \in O_j$. We present the resulting catchment areas in the left panel of Figure 3. In the second case, we increase the costs of the location in the upper left corner by setting $B_o/W_o = 0.85$, and we reduce the costs of the central and lower right corner locations by letting $B_o/W_o = 1.2$. This case is presented in the right panel of Figure 3.

As Figure 3 shows, when production costs are constant across production locations, catchment areas are all equally sized, are all convex, and are all polygons. However, when production costs vary, the catchment areas can take different shapes and sizes, can be non-convex, and are not polygons. Note the resulting complexity in the catchment areas in the example where we add some heterogeneity. Not only do lower cost regions catchment areas' grow, but the characteristics of their neighbors and their location in the square matter as well.

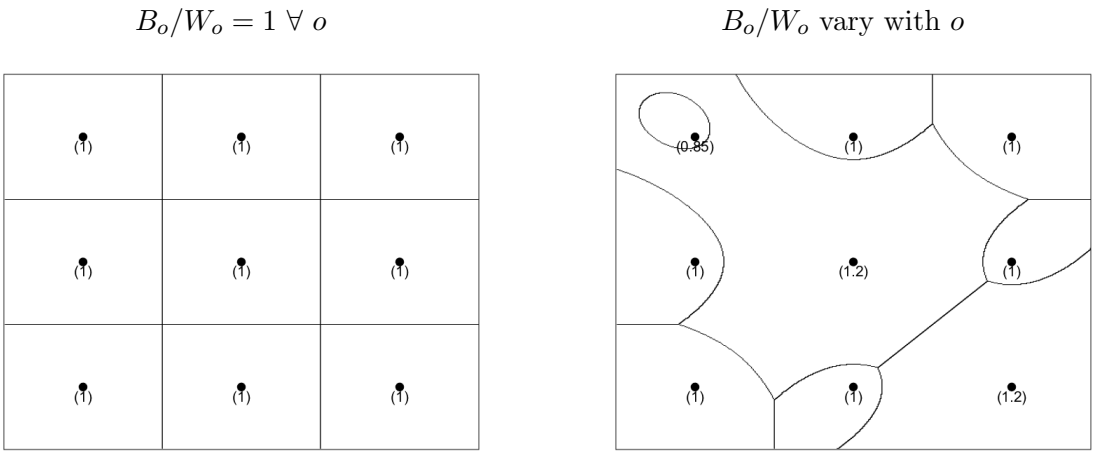
The complexity of this problem seems daunting, particularly once we introduce more locations, richer heterogeneity, and incorporate the decision of how many plants to use and where to locate them. To make progress, we now propose a reformulation of this problem that can be tractably studied, while still preserving its main features and trade-offs.

[5] This problem is equivalent to constructing a weighted Voronoi diagram.

[6] We let plant locations be given by

$$(1/6, 1/6), (1/6, 1/2), (1/6, 5/6), (1/2, 1/6), (1/2, 1/2), (1/2, 5/6), (5/6, 1/6), (5/6, 1/2), (5/6, 5/6) .$$

Figure 3: Catchment Areas



Notes: The figure present the catchment areas, as defined in (2), for the case where 9 plants are located in the squared space. Each dot in each plot corresponds to the location of a plant. Each number on parenthesis next to each dot corresponds to the value for B_o/W_o for that dot.

2.2 A Tractable Limit

We propose a tractable limit of the firm's problem in which the number of plants per firm grows large so that the firm is essentially choosing a density of plants, rather than a discrete number. In particular, we study a limit in which the space that plants take up grows small, trade costs grow large, and the productivity cost for having many plants grows small. We take limits at carefully chosen rates so the problem is well-behaved in the limit. Specifically, for some $\Delta > 0$ let

$$\begin{aligned} \xi^{\Delta} &= \Delta^2, \\ T^{\Delta}(\delta) &= t\left(\frac{\delta}{\Delta}\right), and \\ Z^{\Delta}(q, N) &= z\left(q, \Delta^2 N\right). \end{aligned}$$

We study the limit as $\Delta \to 0$.

We want to study a limit in which the key trade-offs between the fixed and managerial span-of-control costs of setting up plants and the cost of reaching consumers remain relevant; a limit in which plants continue to potentially cannibalize each other's customers. Thus, as Δ declines and the cost of adding plants falls, we increase transport costs. ξ^{Δ} and Z^{Δ} depend on the square of Δ since space is two dimensional while, in contrast, distance is one dimensional.[7] The following proposition describes the profits of the firm in this

[7]There is a natural analogy to the continuous time limit of discrete time portfolio choice problems. In those models, as the

limit.

Proposition 1 *Suppose R_s, D_s, and B_s/W_s are continuous functions of s. Then, in the limit as $\Delta \to 0$, the profits of firm j satisfy*

$$\pi_j = \sup_{n:\mathcal{S}\to\mathbb{R}^+} \int_s \left[x_s z \left(q_j, \int n_{\tilde{s}} d\tilde{s} \right)^{\varepsilon-1} n_s g\left(1/n_s\right) - R_s n_s \right] ds$$

where $x_s \equiv \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^\varepsilon} D_s \left(B_s/W_s\right)^{\varepsilon-1}$ and where $g(u)$ is the integral of $t\left(\cdot\right)^{1-\varepsilon}$ over the distances of points to the center of a regular hexagon with area u.

Proposition 1 shows that in the limit, the firm's problem is one of calculus of variations which, as we show below, is much easier to analyze. The variable x_s combines both local demand facing the plant (D_s) and local cost of effective labor (W_s/B_s) into a measure of local profitability. Before presenting a sketch of the proof of Proposition 1, we discuss a simpler case with identical locations. We then proceed to sketch the proof of Proposition 1 and characterize the solution to the profit maximization problem. The formal proof is relegated to the Appendix.

2.3 Uniform Space

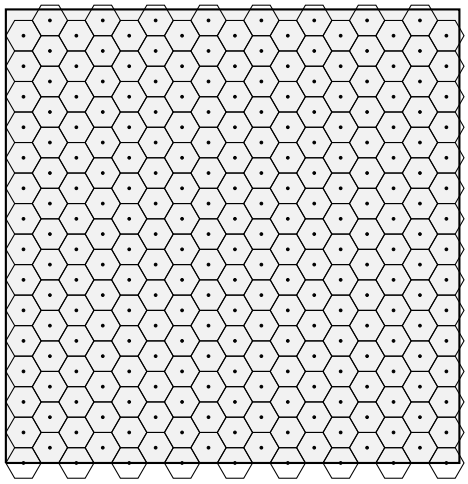
We begin by discussing a simpler case, where all locations are identical. Assume the local profitability of all locations is x and commercial land rents are R. In this special case, the solution to the firm's problem is known. If a firm places N plants, the firm's payoff will be no higher than $xz(q_j, N)^{\varepsilon-1} N_s g\left(|\mathcal{S}|/N\right) - RN$ where $|\mathcal{S}|$ is the area of $\mathcal{S}$ and, as stated in Proposition 1, $g(x)$ is the integral of the transportation cost over a regular hexagon with area x centered at the origin. This follows from the Sum of Moments theorem in Fejes Toth (1953), one of the landmark results in discrete geometry.[8] That is, the nearly optimal policy is to have uniform catchment areas in the form of hexagons, with plants at the center of each hexagon. Figure 4 shows an example of this solution.

Why are hexagonal catchment areas optimal? Jensen's inequality implies that it is optimal to have catchment areas of roughly the same area. Furthermore, optimality dictates that the shape of each catchment area should minimize the average distance from the center to the points in the catchment area. Among all shapes, a circle minimizes this average distance. However, one cannot form a tessellation with circles as they would either overlap or leave empty spots. Among all polygons with which one can construct a uniform tessellation, the hexagon is closest to a circle.

length of a period shrinks to zero, the amount of risk must grow without bound so that there is a meaningful amount of risk to compare across assets. The key, as in our setup, is that the speed at which risk grows is the same as the speed at which the period shrinks. Thus, the relative importance of risk per period unit remains constant. In the appropriate limit, the value of assets follows a Brownian motion.

[8] While the appearance of hexagons as a result of the optimal configuration of economic activity in space is sometimes associated with Christaller (1933), the formal statement and proof are due to Fejes Toth (1953).

Figure 4: Hexagons in a Square



Notes: The figure shows a square of area $|\mathcal{S}|$ divided by N hexagons of area $|\mathcal{S}|/N$.

Note that this is an upper bound. As Figure 4 shows, N disjoint uniform hexagons of size $|\mathcal{S}|/N$ generically do not fit exactly in the space $\mathcal{S}$.[9] It is straightforward to show that if N is large, i.e., if $|\mathcal{S}|$ is large relative to the size of the catchment areas, then the boundary issues is quantitatively less relevant. In the "appropriate" limit, the upper bound is attained.

2.4 Heterogeneous Space

We are interested in understanding the location of a firm's plants in *heterogeneous* space. Proposition 1 provides our key result for this case. The proposition establishes that we can use a similar "large N" limit to obtain a simple characterization of the firm's optimization problem when space is heterogeneous. It states that, in the limit, the optimal policy is to place plants so that local catchment areas are uniform, infinitesimal, hexagons. The variable n_s is the measure of plants in the neighborhood of s, so that $1/n_s$ is a measure of the size of the catchment areas.

2.4.1 A Sketch of the Proof of Proposition 1

Define $b_s \equiv \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}} \left(B_s/W_s\right)^{\varepsilon-1}$. Recall that a firm j's profits are given by

$$\pi_j = \max_{O_j} \left\{ Z\left(q_j, N_j\right)^{\varepsilon-1} \int_s D_s \max_{o \in O_j} \left\{ b_o T \left(\delta_{so}\right)^{1-\varepsilon} \right\} ds - \sum_{o \in O_j} R_o \xi \right\}.$$

[9] Bollobas (1973) showed that the upper bound can be attained only if $\mathcal{S}$ is the union of N disjoint regular hexagons.

We start by dividing the unit square $\mathcal{S}$ into congruent squares with side length k, indexed by $i \in I^k$, denoted by $\mathcal{S}_i^k$ (for any k such that $1/k$ is an integer). For each k, for each square, let $\bar{R}_i^k = \sup_{s \in \mathcal{S}_i^k} R_s$ and $\underline{R}_i^k = \inf_{s \in \mathcal{S}_i^k} R_s$ be the highest and lowest rents in square i. Similarly, define $\bar{D}_i^k = \sup_{s \in \mathcal{S}_i^k} D_s$ and $\underline{D}_i^k = \inf_{s \in \mathcal{S}_i^k} D_s$, $\bar{b}_i^k = \sup_{s \in \mathcal{S}_i^k} b_s$ and $\underline{b}_i^k = \inf_{s \in \mathcal{S}_i^k} b_s$. With these bounds for rents, demand, and the costs per efficiency unit, within these intervals. we can construct upper and lower bounds for profits, $\bar{\pi}_j^{k\Delta}$ and $\underline{\pi}_j^{k\Delta}$. We can then show that

$$\underline{\pi}_j^{k\Delta} \leq \pi_j^{\Delta} \leq \bar{\pi}_j^{k\Delta}.$$

Here, $\bar{\pi}_j^{k\Delta}$ is an upper bound constructed by replacing the rent in each location in square $\mathcal{S}_i^k$ with $\underline{R}_i^k$ and the effective demand and effective productivity in each location in square $\mathcal{S}_i^k$ with $\bar{D}_i^k$ and $\bar{b}_i^k$.[10] Similarly, $\underline{\pi}_j^{k\Delta}$ constructs a lower bound by replacing each R_s, D_s, b_s with $\bar{R}_i^k$, $\underline{D}_i^k$, and $\underline{b}_i^k$ for the locations in square, $\mathcal{S}_i^k$. In constructing the lower bound, we impose the ad hoc restriction that each plant located in $\mathcal{S}_i^k$ only sells to households in $\mathcal{S}_i^k$, and all plants within $\mathcal{S}_i^k$ have catchment areas that are formed by regular hexagons; since regular hexagons do not form a tessellation of a square, not all customers in $\mathcal{S}_i^k$ are served by the firm in this suboptimal policy.

The second step is to study the limit as $\Delta \to 0$. Define the function $\kappa(n_s) \equiv n_s g(1/n_s)$. Then, for any k, we prove that $\bar{\pi}_j^k$ and $\underline{\pi}_j^k$ are such that

$$\lim_{\Delta \to 0} \bar{\pi}_j^{k\Delta} \leq \bar{\pi}_j^k \equiv \sup_{\{n_s \geq 0\}} \int \left\{ \bar{D}_s^k \bar{b}_s^k z \left(q_j, \int n_{\tilde{s}} d\tilde{s} \right)^{\varepsilon - 1} \kappa(n_s) - n_s \underline{R}_s^k \right\} ds,$$

and

$$\lim_{\Delta \to 0} \underline{\pi}_j^{k\Delta} \geq \underline{\pi}_j^k \equiv \sup_{\{n_s \geq 0\}} \int_s \left\{ \underline{D}_s^k \underline{b}_s^k z \left(q_j, \int n_{\tilde{s}} d\tilde{s} \right)^{\varepsilon - 1} \kappa(n_s) - n_s \bar{R}_s^k \right\} ds,$$

which imply that

$$\underline{\pi}_j^k \leq \pi_j \leq \bar{\pi}_j^k,$$

where $\pi_j \equiv \lim_{\Delta \to 0} \pi_j^{\Delta}$. For any k, the economic features are uniform within each square $\mathcal{S}_i^k$, so we can use results from discrete geometry to derive relatively simple expressions for the bounds. For the upper bound, the results imply that the catchment areas within each square $\mathcal{S}_i^k$ form a mesh with uniform regular hexagons. For the lower bound, the ad hoc restriction imposes that the catchment areas form a mesh with uniform regular hexagons.

The final step is to show that

$$\lim_{k \to 0} \bar{\pi}_j^k = \lim_{k \to 0} \underline{\pi}_j^k = \sup_{n_s \geq 0} \int_s \left\{ D_s b_s z \left(q_j, \int n_{\tilde{s}} d\tilde{s} \right)^{\varepsilon - 1} \kappa(n_s) - R_s n_s \right\} ds,$$

[10] In constructing this upper bound, we compute separately for each $k \times k$ square, the profits the firm would earn from the plants in $\mathcal{S}_i^k$, if it had no other plants in all of $\mathcal{S}$. We then add together the profits for all $k \times k$ squares. Clearly, we are constructing an upper bound for profits, not the value of profit for any feasible policy. Note also that, when we take the limit $\Delta \to 0$, sales outside a plant's $k \times k$ square go to zero.

i.e., that the limit can be interchanged with the supremum. Since $x_s = D_s b_s$, this proves Proposition 1. The rest of the technical details of the proof are relegated to the Appendix.

2.4.2 The Local Efficiency of Distribution and its Properties

The function $\kappa(n_s) \equiv n_s g(1/n_s)$ can be interpreted as the *local efficiency of distribution* in the neighborhood of s. It represents the fraction of the value of sales a firm retains after subtracting the cost of optimally transporting the goods to consumers from its n_s plants. Figure 5 provides a one dimensional representation of the function $g(\cdot)$ and the implied function $\kappa(\cdot)$. The following lemma describes some useful properties of κ.

Lemma 2 *$\kappa(n) \equiv ng\left(\frac{1}{n}\right)$ is strictly increasing and strictly concave, and satisfies the following properties:*

1. $\kappa(0) = 0$;
2. $\lim_{n\to\infty} \kappa(n) = 1$;
3. $1 - \kappa(n) \underset{n\to\infty}{\sim} n^{-1/2}$

If transport costs satisfy $\lim_{\delta\to\infty} \delta^{-\frac{4}{\varepsilon-1}} t(\delta) = \infty$ then

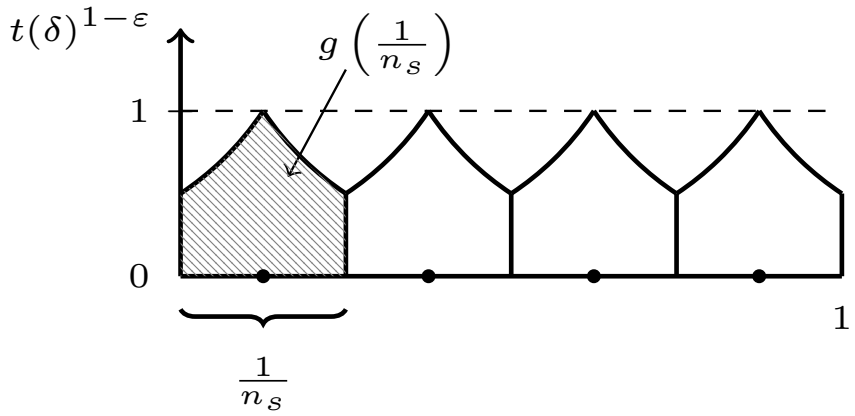
4. $\kappa''(0) = 0$;
5. $\kappa'(0) = \int_0^\infty t(\delta)^{1-\varepsilon} 2\pi\delta d\delta < \infty$.

The first property says that without plants revenues are zero. The second property states that as n grows infinitely large, catchment areas grow small and $\kappa(n)$ approaches an upper bound of 1. Additional plants only cannibalize other plants without significant gains from reducing transport costs; the economy becomes "saturated". The third property states that $\kappa(n)$ follows an asymptotic power law as n grows large. If, asymptotically, trade costs increase sufficiently fast with distance, we can give a sharper characterization of the efficiency of distribution when n is small. The fourth property states that when the number of plants is small, local profits increase linearly in the number of plants. In sum, cannibalization is irrelevant for the first set of plants but the dominant force when the number of plants grows large. Finally, the fifth property, says that there is no Inada condition at $n = 0$. Hence, there can be locations s in which the firm places no plants, $n_s = 0$.

2.5 The Assignment of Plants to Locations

Proposition 1 can be used to characterized how firms place their plants. As before, we assume that the firm takes as given the distribution of commercial rents, R_s, and the distribution of local profitability, x_s. The

Figure 5: A One Dimensional Representation of the Efficiency of Distribution, $\kappa(n_s) \equiv n_s g(1/n_s)$



problem of choosing how many plants to have, N_j, and their distribution in space, $n_j : \mathcal{S} \to \mathbb{R}^+$, can be stated as

$$\sup_{N_j, n_j : \mathcal{S} \to \mathbb{R}^+} \int_s \left[x_s z(q_j, N_j)^{\varepsilon - 1} \kappa(n_{js}) - R_s n_{js} \right] ds,$$

subject to

$$\int_s n_{js} ds \leq N_j.$$

Letting λ_j be the multiplier on the constraint, the first order condition with respect to n_{js} is given by

$$x_s z_j^{\varepsilon - 1} \kappa'(n_{js}) \leq R_s + \lambda_j, \qquad \text{with equality if } n_{js} > 0, \tag{3}$$

where we use z_j as shorthand for $z(q_j, N_j)$. The first order condition with respect to N_j is

$$\lambda_j = -\frac{d[z(q_j, N_j)^{\varepsilon - 1}]}{dN_j} \int_s x_s \kappa(n_{js}) ds. \tag{4}$$

To characterize the solution to this problem it is useful to make the following assumption on the productivity function $z(q, N)$.

Assumption 2 *$z(q, N) = q\Xi(N)$, where Ξ is a log-concave function.*

Using the first order conditions in (3), together with this assumption, we can show that firms with higher endogenous productivity have higher marginal cost of increasing the number of plants, even relative to their firm-specific profitability, $z_j^{\varepsilon - 1}$.

Lemma 3 *Consider two firms with $z_1 < z_2$. Then, either $\frac{\lambda_1}{z_1^{\varepsilon-1}} < \frac{\lambda_2}{z_2^{\varepsilon-1}}$ or $N_1 = N_2 = 0$.*

Proof. Since κ is concave, the density of plants is a decreasing function of $\frac{R_s+\lambda_j}{x_s z_j^{\varepsilon-1}}$. Suppose that $\frac{\lambda_1}{z_1^{\varepsilon-1}} \geq \frac{\lambda_2}{z_2^{\varepsilon-1}}$. Then, in every market $\frac{R_s+\lambda_1}{x_s z_1^{\varepsilon-1}} > \frac{R_s+\lambda_2}{x_s z_2^{\varepsilon-1}}$. Therefore $n_{1s} \leq n_{2s}$ with a strict inequality whenever $n_{2s} > 0$. If $N_2 > 0$, then $N_2 > N_1$, and the log-concavity of z with respect to N along with $\kappa' > 0$ implies $\frac{\lambda_1}{z_1^{\varepsilon-1}} = (\varepsilon-1)\frac{-z_N(q_1,N_1)}{z(q_1,N_1)}\int_s x_s \kappa(n_{1s})ds < (\varepsilon-1)\frac{-z_N(q_2,N_2)}{z(q_2,N_2)}\int_s x_s \kappa(n_{2s})\,ds = \frac{\lambda_2}{z_2^{\varepsilon-1}}$, a contradiction. If $N_2 = 0$, then $N_1 = 0$. ■

Our next result uses the previous Lemma to prove that more productive firms set up relatively more plants in locations with higher rents.

Proposition 4 *Consider two firms with $z_1 < z_2$. Let $R^*(z_1, z_2)$ be the unique rent that satisfies*

$$\frac{R^*(z_1,z_2)+\lambda_2}{R^*(z_1,z_2)+\lambda_1} = \frac{z_2^{\varepsilon-1}}{z_1^{\varepsilon-1}}.$$

Then, $R_s > R^*(z_1,z_2)$ implies that $n_{2s} \geq n_{1s}$, with strict inequality if $n_{2s} > 0$; $R_s < R^*(z_1,z_2)$ implies that $n_{1s} \geq n_{2s}$, with strict inequality if $n_{1s} > 0$; and $R_s = R^*(z_1,z_2)$ implies that $n_{1s} = n_{2s}$.

Proof. $z_2 > z_1$ implies that $\lambda_2 > \lambda_1$ and $\frac{\lambda_2}{z_2^{\varepsilon-1}} > \frac{\lambda_1}{z_1^{\varepsilon-1}}$. Therefore $\frac{R+\lambda_2}{R+\lambda_1} > 1$, so $\frac{z_1^{\varepsilon-1}}{z_2^{\varepsilon-1}}\frac{R+\lambda_2}{R+\lambda_1}$ is strictly decreasing in R. Since $\lim_{R\to 0}\frac{z_1^{\varepsilon-1}}{z_2^{\varepsilon-1}}\frac{R+\lambda_2}{R+\lambda_1} = \frac{\lambda_2/z_2^{\varepsilon-1}}{\lambda_1/z_1^{\varepsilon-1}} > 1$ and $\lim_{R\to\infty}\frac{z_1^{\varepsilon-1}}{z_2^{\varepsilon-1}}\frac{R+\lambda_2}{R+\lambda_1} = \frac{z_1^{\varepsilon-1}}{z_2^{\varepsilon-1}} < 1$, there is a unique R^* such that $\frac{z_1^{\varepsilon-1}}{z_2^{\varepsilon-1}}\frac{R+\lambda_2}{R+\lambda_1} = 1$. If $R_s > R^*(z_1,z_2)$ and $n_{2s}, n_{1s} > 0$ then $\kappa'(n_{2s}) = \frac{R_s+\lambda_2}{z_2^{\varepsilon-1}x_s} < \frac{R_s+\lambda_1}{z_1^{\varepsilon-1}x_s} = \kappa'(n_{1s})$ and since κ' is decreasing, $n_{2s} > n_{1s}$. If $n_{2s} > 0$ and $n_{1s} = 0$, then of course $n_{2s} > n_{1s}$. If $n_{2s} = 0$, then $\kappa'(0) \leq \frac{R_s+\lambda_2}{z_2^{\varepsilon-1}x_s} < \frac{R_s+\lambda_1}{z_1^{\varepsilon-1}x_s}$, which implies that it is optimal for $n_{1s} = 0$. The argument for $R_s < R^*(z_1,z_2)$ is identical. The argument for $R = R^*(z_1,z_2)$ is trivial. ■

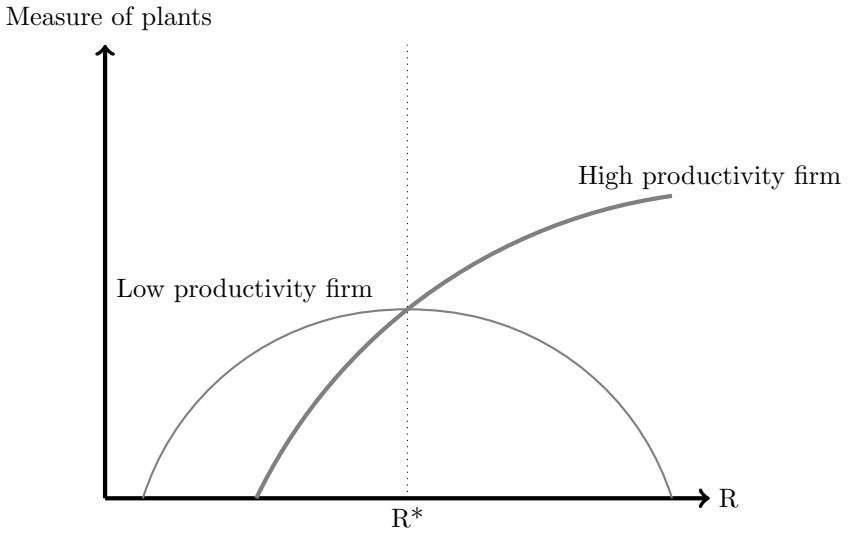
Proposition 4 states that for two firms with different productivities, there is a cutoff level of rent such that the firm with higher productivity places more plants in locations with higher rent and the firm with lower productivity places more plants in locations with lower rent. Thus, even while the two firms have overlapping footprints, there is a clear pattern of sorting. Figure 6 provides a graphical representation of this result.

The type of sorting implied by Proposition 4 stands in sharp contrast to workhorse models of trade and multinational production in which the more marginal locations are reached by the most productive firms.[11] Here, it is the less productive firms that go to the lower rent locations. Why the difference?

Firms balance the marginal profit from an additional plant, $x_s z_j^{\varepsilon-1}\kappa'(n_{js})$, against the effective fixed cost of a new plant, $R_s + \lambda_j$, which depends on the local rent and the productivity penalty arising from the larger span of control of managers. As Lemma 3 shows, higher productivity firms have higher span of control costs, as given by λ_j, relative to their profitability, $z_j^{\varepsilon-1}$. Hence, a higher rent has a smaller impact on their costs. Formally, since $\frac{d\ln(R_s+\lambda_j)}{d\ln R_s}$ is decreasing in λ_j, the effective fixed cost of setting up a plant rises proportionally less with rents for the high productivity firm, and so it places relatively more plants in

[11] See for example Melitz (2003), Eaton and Kortum (2002), and Ramondo and Rodríguez-Clare (2013).

Figure 6: Location of Plants of a High and a Low Productivity Firm



high rent locations. Most models of plant location decisions in the literature do not feature cannibalization and span of control costs, and so they miss this implication.

We can extend the result further to show that the marginal efficiency of distribution of more productive firms is relatively smaller in higher rent locations.

Lemma 5 *Consider two firms with $z_1 < z_2$ and two locations with $R_s < R_{\hat{s}}$. Then, if n_{1s}, $n_{1\hat{s}}$, n_{2s}, $n_{2\hat{s}} > 0$,*

$$\frac{\kappa'(n_{2s})}{\kappa'(n_{1s})} > \frac{\kappa'(n_{2\hat{s}})}{\kappa'(n_{1\hat{s}})}.$$

Proof. Since $\lambda_2 > \lambda_1$, $\frac{R+\lambda_2}{R+\lambda_1} > 1$, so $\frac{R+\lambda_2}{R+\lambda_1}$ is decreasing in R. Hence, we have that

$$\frac{\kappa'(n_{2s})}{\kappa'(n_{1s})} = \frac{z_1^{\varepsilon-1}}{z_2^{\varepsilon-1}} \frac{R_s + \lambda_2}{R_s + \lambda_1} > \frac{z_1^{\varepsilon-1}}{z_2^{\varepsilon-1}} \frac{R_{\hat{s}} + \lambda_2}{R_{\hat{s}} + \lambda_1} = \frac{\kappa'(n_{2\hat{s}})}{\kappa'(n_{1\hat{s}})}.$$

■

So far we have shown that firms sort based on local commercial rents. Similarly, we can show that firms sort according to local profitability, x_s. The next proposition shows that, if there is sorting, high productivity firms set relatively more plants in locations with high profitability.

Proposition 6 *Consider two firms with $z_1 < z_2$. Suppose there is a location s^* such that $n_{1s^*} = n_{2s^*} = n^* > 0$. Then, for any location s with $x_s > x^*$, $n_{2s} > n_{1s}$, and for any location s with $x_s < x^*$, $n_{2s} < n_{1s}$.*

Proof. Suppose that $x_s > x^*$. The first order condition $z_j^{\varepsilon-1}\kappa'(n_{js}) = \frac{R_s+\lambda_j}{x_s}$ implies that

$$z_2^{\varepsilon-1}\left[\kappa'(n_{2s})\, x_s - \kappa'(n^*)\, x_{s^*}\right] = R_s - R_{s^*} = z_1^{\varepsilon-1}\left[\kappa'(n_{1s})\, x_s - \kappa'(n^*)\, x_{s^*}\right],$$

which can be rearranged as

$$1 > \frac{x^*}{x_s} = \frac{z_2^{\varepsilon-1}\kappa'(n_{2s}) - z_1^{\varepsilon-1}\kappa'(n_{1s})}{z_2^{\varepsilon-1}\kappa'(n^*) - z_1^{\varepsilon-1}\kappa'(n^*)} = \frac{z_2^{\varepsilon-1}\frac{\kappa'(n_{2s})}{\kappa'(n^*)} - z_1^{\varepsilon-1}\frac{\kappa'(n_{1s})}{\kappa'(n^*)}}{z_2^{\varepsilon-1} - z_1^{\varepsilon-1}}.$$

Since $z_1 < z_2$, the inequality implies that $\kappa'(n_{2s}) < \kappa'(n_{1s})$, or $n_{2s} > n_{1s}$. The $x_s < x^*$ case can be proven analogously. ■

The results above condition on firms with a positive density of plants in particular locations. Our next result shows that, for any given location, there is a productivity threshold such that firms with productivity below the threshold do not set up plants in that location. Under further restrictions on the span of control costs, the lemma also shows that there is another threshold such that firms with high enough productivity do not set up plants there either. That is, when all conditions are satisfied, only plants with productivities between these thresholds set up plants in a given location.

Lemma 7 *If* $\lim_{\delta\to\infty} \frac{\delta^2}{t(\delta)^{\varepsilon-1}} = 0$*, for any location* s*, there exists a productivity threshold* $\underline{z}_s > 0$ *such that* $n_{js} = 0$ *if* $z_j < \underline{z}_s$*. If* $\lim_{z\to\infty} \frac{\lambda_j}{z^{\varepsilon-1}} = \infty$*, then there exists an additional threshold* $\bar{z}_s < \infty$ *such that* $n_{js} = 0$ *if* $z_j > \bar{z}_s$*.*

Proof. $n_s(z)$ denotes the density of plants a firm with productivity z places in location s. The first order condition (3) implies that $n_s(z) = 0$ if $\frac{R_s+\lambda_j}{x_s z^{\varepsilon-1}} > \kappa'(0)$. Since $\lambda_j > 0$,

$$\lim_{z\to 0} \frac{R+\lambda_j}{xz^{\varepsilon-1}} \geq \lim_{z\to 0} \frac{R}{xz^{\varepsilon-1}} = \infty,$$

and, if $\lim_{z\to\infty} \frac{\lambda_j}{z^{\varepsilon-1}} = \infty$,

$$\lim_{z\to\infty} \frac{R+\lambda_j}{xz^{\varepsilon-1}} = \frac{1}{x} \lim_{z\to\infty} \frac{\lambda_j}{z^{\varepsilon-1}} = \infty.$$

The result follows from the fact that κ' is continuous, strictly decreasing, and $\kappa'(0) < \infty$ if $\lim_{\delta\to\infty} \frac{\delta^2}{t(\delta)^{\varepsilon-1}} = 0$ by Lemma 2. ■

Our final result in this subsection refers to the total size of firms. The results above condition on a firm's productivity. However, empirically, it is easier to condition on other firm observables, such as their total employment size or the total number of plants. We do not have a result that the total number of plants is increasing in firm productivity. Not only do firms sort their plants across locations, but their optimal plant size varies depending on local characteristics. However, under particular parametric assumptions on a firm's productivity function, and if wages are constant across space, we can show that more productive firms employ more workers.[12] We let L_j denote the total number of workers of firm j and l_{js} denote its density of employment in location s.

Lemma 8 *Suppose that* $z(q, N) = qe^{-\zeta N}$ *and local wages are constant across locations at* W*. Consider two firms with* $z_1 < z_2$*, then either* $L_1 < L_2$ *or* $L_1 = L_2 = 0$*.*

[12] The assumption of equal local wages is consistent with the general equilibrium framework we setup in Section 3.

Proof. For firm j, variable profit in location s is $x_s z_j^{\varepsilon-1}\kappa(n_{js})$, so with a markup of $\frac{\varepsilon}{\varepsilon-1}$, the expenditure on labor in s is $W_s l_{js} = (\varepsilon-1)\, x_s z_j^{\varepsilon-1}\kappa(n_{js})$. Since the wage $W_s = W$ for all $s \in \mathcal{S}$, j's total employment is $L_j = \frac{\varepsilon-1}{W}\int x_s z_j^{\varepsilon-1}\kappa(n_{js})\,ds = \frac{1}{W}\frac{1}{\zeta}\lambda_j$. If $N_1 = N_2 = 0$, then $L_1 = L_2 = 0$. Otherwise, by Lemma 3, $\frac{\lambda_2}{z_2^{\varepsilon-1}} > \frac{\lambda_1}{z_1^{\varepsilon-1}}$, which implies $\lambda_2 > \lambda_1$, and so $L_2 > L_1$. ■

3 Equilibrium

We now proceed to embed the problem of the multi-plant firm that we studied in the previous section into a general equilibrium framework. The purpose is to illustrate how the proposed limit of the firm's problem can be readily incorporated in a quantitative spatial model. We choose a particular equilibrium model to do so, but many alternative general equilibrium setups would work as well. We start by embedding the standard problem in our equilibrium setup and then characterize the limit when $\Delta \to 0$. We let locations in $\mathcal{S}$ be heterogeneous in their exogenous productivity, B_s, and an exogenous amenity, A_s. Denote the endogenous number of workers at s by $\mathcal{L}_s$.

3.1 Workers

The economy consists of a mass $\mathcal{L}$ of workers who can freely move across space. They choose a location where to live and work, and supply one unit of labor inelastically in that location. They consume a consumption bundle, c, housing space, h, and amenities according to the utility function

$$u(c, h, A) = Ac^{1-\eta}h^{\eta}. \tag{5}$$

Consistent with Assumption 1, c is a Dixit-Stiglitz bundle of all varieties j, with elasticity of substitution across varieties ε. A worker in location s earns a wage W_s, faces price of consumption bundle P_s, and faces rental cost of housing R_s^H per unit of housing space. In addition, the household owns a share of a mutual fund that owns all firms and all land, and receives from the mutual fund a lump sum transfer Υ. Hence, the budget constraint of the worker is given by

$$P_s c + R_s^H h \leq W_s + \Upsilon. \tag{6}$$

Then, if c_s denotes the optimal choice of the CES consumption bundle by agents in s, $D_s(p_{js}) = \mathcal{L}_s c_s P_s^{\varepsilon} p_{js}^{-\varepsilon}$ is the residual demand function.[13] Hence, the local demand level can be defined as $D_s = \mathcal{L}_s c_s P_s^{\varepsilon}$.

3.2 Land Use

There is a unit measure of land in each location that is owned by a national mutual fund. Competitive developers can rent the land at price R_s^M and use it for either commercial real estate, renting it to plants

[13] Where we are already using the implication that identical agents make identical choices in each location.

at rate R_s, or residential housing, renting it to workers at rate R_s^H. Housing and commercial real estate are perfect substitutes and satisfy the constraint

$$H_s + \mathcal{N}_s \leq 1, \tag{7}$$

where H_s is total measure of housing, $\mathcal{N}_s$ is the total measure of commercial real estate. Note that $\mathcal{N}_s/\xi$ is the measure plants at location s. In addition, since all workers in the same location choose the same amount of housing, $h_s = H_s/L_s$.

3.3 Equilibrium definition and limit economy

We are ready to define a competitive equilibrium.

Definition 1 *Given a set of local fundamentals, $\{A_s, B_s\}_{s\in\mathcal{S}}$, an equilibrium is a population, land, and consumption allocation $\{\mathcal{L}_s, \mathcal{N}_s, h_s, c_s\{c_{js}\}_{j\in J}\}_{s\in\mathcal{S}}$, a set of prices $\{W_s, R_s, R_s^M, R_s^H, P_s\}_{s\in\mathcal{S}}$, a set of firm choices, O_j, $\{l_{jo}, S_{jo}\}_{o\in O_j}$, $\{p_{js}\}_{s\in\mathcal{S}}$, and profits π_j, for each j, a level of utility $\bar{u}$ and a transfer payment Υ, such that:*

- *Individual maximize (5) subject to (6) and are indifferent across locations for which there is positive employment, so $\bar{u} \geq u(c_s, h_s, A_s)\forall s \in \mathcal{S}$, with equality if $\mathcal{L}_s > 0$.*
- *The price and consumption indexes satisfy $P_s^{1-\varepsilon} = \int_j p_{js}^{1-\varepsilon} dj$ and $c_s^{\frac{\varepsilon-1}{\varepsilon}} = \int_j c_{js}^{\frac{\varepsilon-1}{\varepsilon}} dj$.*
- *Developers maximize profit. Hence, given rents, $\{H_s, \mathcal{N}_s\} = \arg\max R_s^H h_s \mathcal{L}_s + R_s \mathcal{N}_s - R_s^M$.*
- *Firms solve the problem in (1) and (2) for each j.*
- *The transfer payment is given by $\Upsilon = \int_s R_s^M ds + \int_j \pi_j dj$.*
- *Goods, labor, and land markets clear. Thus, for each j and $o \in O_j$,*

$$l_{jo} = \int_{s\in CT(o)_j} \mathcal{L}_s c_s P_s^{\varepsilon} p_{js}^{-\varepsilon} \frac{T(\delta_{so})}{B_o Z(q_j, N_j)},$$

$\mathcal{L} = \int_s \mathcal{L}_s ds$, and, $\forall s \in \mathcal{S}$, $\mathcal{L}_s = \int_j l_{js} I_{\{s\in O_j\}} dj$ and (7) hold.[14]

In the Appendix we prove that, when $\Delta \to 0$, the equilibrium defined above has a standard structure. Using only the decisions of firms and the definition of the Dixit-Stiglitz consumption aggregate, we can characterize a location's aggregate productivity in Proposition 9.

Proposition 9 *In the limit when $\Delta \to 0$, the number of plants is given by $\mathcal{N}_s = \int_j n_{js} dj$, and the local price index and consumption bundle satisfy $P_s = \frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s \mathcal{Z}_s}$, and $c_s = B_s \mathcal{Z}_s$, where $\mathcal{Z}_s \equiv \left(\int_j z_j^{\varepsilon-1} \kappa(n_{js}) dj\right)^{\frac{1}{\varepsilon-1}}$.*

[14] Where $I_{\{s\in O_j\}}$ denotes an indicator function equal to 1 when $s \in O_j$ and 0 otherwise.

Observe that aggregate productivity in a location, $\mathcal{Z}_s$, is the CES aggregate of the firms' effective productivities, z_j, with weights given by the efficiency of distribution, $\kappa(n_{js})$. Then, the local price index is just the standard CES markup times the 'aggregate' local marginal cost. Similarly, the local consumption bundle is simply the local labor productivity. With Proposition 9 in hand, we can solve easily for the rest of the equilibrium. The local profitability is given by $x_s = \frac{1}{\varepsilon-1}\frac{W_s\mathcal{L}_s}{\mathcal{Z}_s^{\varepsilon-1}}$. The utility specification in (5) implies that $P_s c_s = (1-\eta)(W_s+\Upsilon)$, and so the nominal wage solves $\frac{\varepsilon}{\varepsilon-1}W_s = (1-\eta)(W_s+\Upsilon)$. Hence, the nominal wage is the same in all locations, and proportional to the transfer payment Υ.

Perfect mobility implies that all workers obtain utility $\bar{u} = A_s c_s^{1-\eta} h_s^{\eta}$. Since $h_s\mathcal{L}_s = H_s$, population in location s is $\mathcal{L}_s = \left[\frac{1}{\bar{u}}A_s B_s^{1-\eta}\mathcal{Z}_s^{1-\eta}H_s\right]^{1/\eta}$, and so the labor market equilibrium condition implies, as is standard in this class of models, that

$$\frac{\mathcal{L}_s}{\mathcal{L}} = \frac{\left[A_s B_s^{1-\eta}\mathcal{Z}_s^{1-\eta}H_s\right]^{1/\eta}}{\int_s \left[A_{\tilde{s}} B_{\tilde{s}}^{1-\eta}\mathcal{Z}_{\tilde{s}}^{1-\eta}H_s\right]^{1/\eta} d\tilde{s}}.$$

Note that developer optimization implies that $R_s = R_s^H = R_s^M$ (since all locations have some establishments and population in the limit). Consumer optimization implies that $R_s h_s = \eta(W_s+\Upsilon)$, which together with zero profits for developers and the land market clearing condition in (7), determine h_s, R_s and R_s^M.

In sum, Proposition 9 gives us the key local aggregates to solve the general equilibrium of the model as in standard spatial equilibrium frameworks.[15] Of course, to do so requires knowing the function κ as well as the solution to the firm's problem which we characterized in Section 2.

3.4 Numerical Illustration of an Industry Equilibrium

To illustrate more concretely some of the equilibrium implications of our theory, we now specify all relevant functional forms and distributions and solve for an equilibrium of the model numerically. Our parametrization is intended to make the relevant forces visually clear and transparent. We study the implications of a single 'small' industry when transport or span of control costs change for that industry only. Hence, in the comparative statics exercises, we keep rents and wages fixed which implies that firms within an industry interact exclusively through the local industry price index.

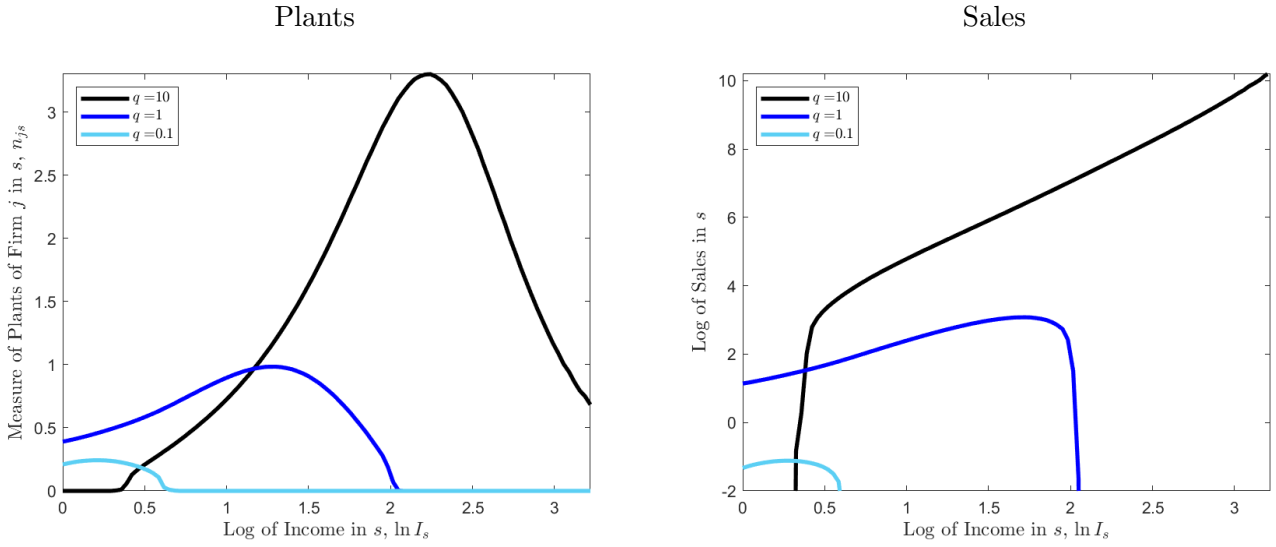
Assume that the economy is formed by a unit continuum of industries, and that households have symmetric Cobb-Douglas preferences across these industries so that expenditure on any mass of industries is a constant fraction of local income. Expenditure in location s, I_s, is distributed truncated Pareto, so that the measure of locations with income weakly less than I is $[1-(I/\underline{I})^{-\chi_I}]/[1-(\bar{I}/\underline{I})^{-\chi_I}]$, with $\underline{I}=1$, $\bar{I}=25$, and $\chi_I = 2$. We set the elasticity of substitution across varieties, ε, to 2. We assume that the amenity distribution is such that the rent schedule in a location with income I_s is given by $R(I_s) = e^{\log(I_s)^2}$.

Let transportation costs take the form $t(\delta;\phi) \equiv t(\delta/\sqrt{\phi})$, where ϕ indexes the efficiency of transportation

[15] See Redding and Rossi-Hansberg (2017) for a review of the quantitative spatial economics literature.

(i.e. a higher ϕ implies lower trade costs for a given distance traveled, δ).[16] This implies that $\kappa(n;\phi) \equiv \kappa(\phi n)$. We parameterize transportation costs as $t(\delta/\sqrt{\phi}) = e^{\delta/\sqrt{\phi}}$. We set $\phi = 0.04$. Firms' productivity is given by $z(q,N) = qe^{-N/\sigma}$, where σ indexes the efficiency of a firm's span of control (i.e. a higher σ implies a higher z for the same aggregate size of the firm, N). We set $\sigma = 1$. There is a unit measure of firms, and the distribution of productivity is given by a truncated Pareto distribution so that the measure of firms with pure productivity no greater than q is $[1-(q/\underline{q})^{-\chi_q}]/[1-(\bar{q}/\underline{q})^{-\chi_q}]$, with $\underline{q} = 0.1$, $\bar{q} = 10$, and $\chi_q = 1.25$.

Figure 7: Sorting in Industry Equilibrium



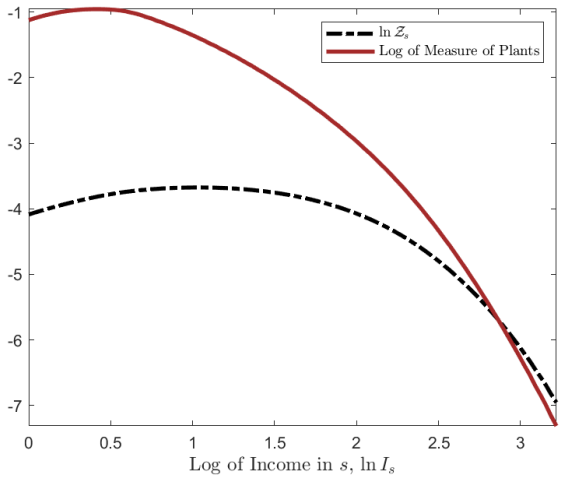
We first describe the baseline industry equilibrium and then proceed to study comparative static exercises with respect to σ and ϕ. Appendix E describes the numerical algorithm that solves for the industry equilibrium. Figure 7 presents the distribution of plants, n_{js}, and sales, $(\varepsilon-1)z_j^{\varepsilon-1}x_s\kappa(n_{js})$, for three representative firms: a firm with the lowest productivity, $q = 0.1$, a firm with intermediate productivity, $q = 1$, and a firm with the highest productivity, $q = 10$. As implied by Proposition 6, for any pair of firms, there exists an income threshold (or, equivalently, a rent threshold since rents are monotone in income) such that the more productive of the two firms sets up more plants above the threshold and fewer plants below. In our example, the most productive firm operates many plants in middle income locations and fewer plants in very high or very low income locations (the case $q = 10$ in the left-hand panel of Figure 7). In fact, it operates no plants in the worse locations. The logic should be clear; rents in high income locations are high which encourages high productivity firms to economize on plants at the cost of having lower efficiency of distribution, $\kappa(n_{js})$. As shown in the right-hand panel of Figure 7, they compensate with higher sales from each plant which results in higher total sales. Low income locations, in contrast, are less attractive to large

[16] Note ϕ is defined so that it enters the function t as $\sqrt{\phi}$, while it enters the function κ linearly. The reason for the discrepancy is that the function κ is constructed from an integral over a two dimensional space.

firms, since their shadow cost of setting up an additional plant is high given the productivity penalty that arises from their larger span of control (λ_j is increasing in q_j). Again, these firms compensate with higher sales from each plant. Firms with lower productivity then take advantage of low income locations given their lower span of control and the lack of competition from top firms in those locations.

The total measure of plants across locations depends on both the set of firms that open plants in each location and how many plants they open. High income locations accommodate only the best firms which operate few plants because of the associated high rents. Low income locations attract the worst firms, which set up few plants since they have low productivity. In general, the overall dependence of the total number of plants on the level of income depends on the distribution of amenities and other characteristics of the economy through the rent function. Figure 8 shows that in this example, the log of the measure of plants is hump-shaped. Figure 8 also presents the log of local industry productivity, $\mathcal{Z}_s$, which is also hump-shaped though its maximum lies to the right of that of the measure of plants. As Proposition 9 shows, local industry productivity depends on the productivity of firms that locate in the region, which is strictly increasing in local income, but also on the local efficiency of distribution, which is lower in high-income locations because of higher rents. Ultimately, however, local profitability for the industry, $x_s = \frac{1}{\varepsilon-1}\frac{I_s}{\mathcal{Z}_s^{\varepsilon-1}}$, remains strictly increasing in I_s.

Figure 8: Log of Total Measure of Plants and Industry Productivity by Location



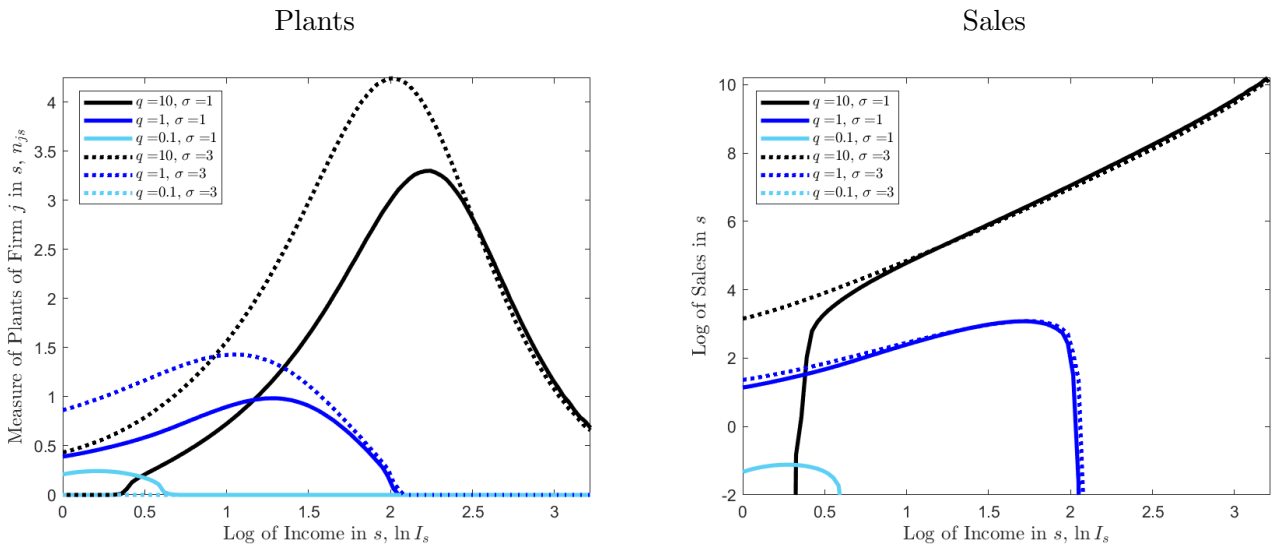
3.4.1 Improvements in an Industry's Span of Control Technology

Consider the effect of an improvement in the span of control technology captured by an increase in the parameter σ in the firm's productivity function, $z(q, N) = qe^{-N/\sigma}$. A better span of control technology increases firm productivity and lowers the shadow cost of adding new plants. This motivates firms to have more plants in more locations. In equilibrium, the additional entry leads to more local competition, through

an increase in $\mathcal{Z}_s$ at all locations, which makes some firms shrink and others exit from some, or all, locations.

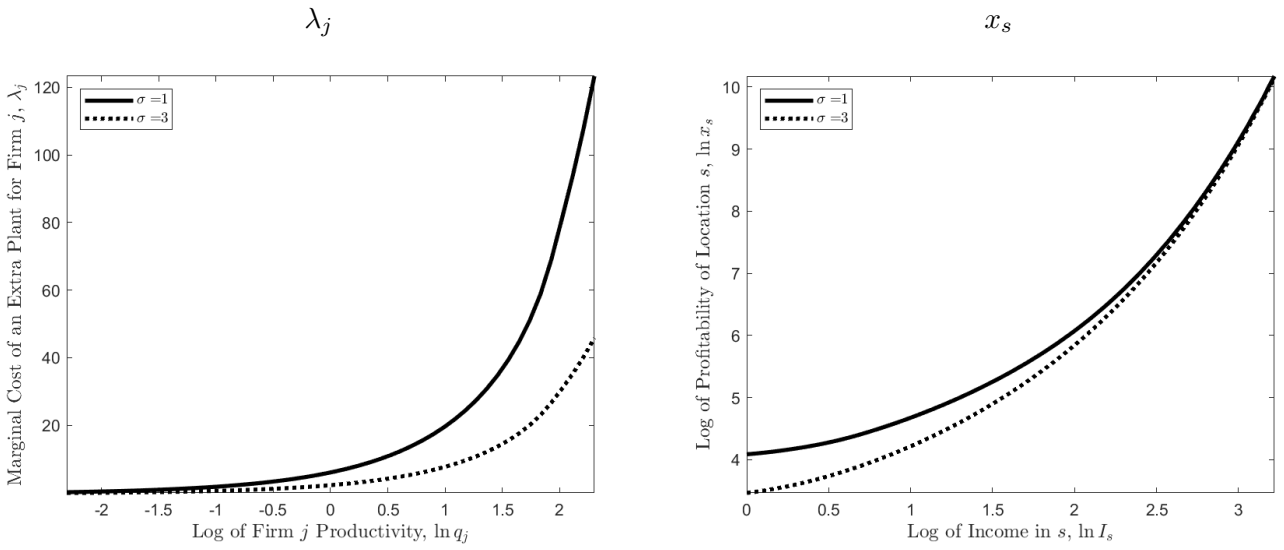
Figure 9 reproduces Figure 7 (the solid lines computed for $\sigma = 1$) and compares it with findings for $\sigma = 3$ (the dashed lines). In response to the improvement in span of control technology, the top firm increases the measure of plants in low-income locations. It also reduces its presence slightly in the highest-income markets because of increased competition. The middle-productivity firm expands its presence in both lower and higher-income locations. Holding fixed the actions of other firms, the lowest productivity firm would benefit from the improved span of control technology as well. However, increased competition pushes it to exit all markets. The top firm not only enters lower-income markets but, with improved span of control technology, ends up outselling the medium productivity firm that already had a presence in those locations. The ability to manage a greater span of control, therefore, results in a net reallocation of sales from low to high-productivity firms.

Figure 9: Span of Control and Sorting in an Industry Equilibrium



The left-hand panel in Figure 10 shows how an improved span of control technology affects the shadow cost of additional plants, λ_j. As argued above, λ_j declines following the direct effect of the technological change. The effect is clearly magnified for high-productivity firms. These firms benefit most since their better technology makes them want to expand more extensively in space, and thus makes them benefit disproportionately from a technology that renders such an expansion less costly. The right-hand panel in Figure 10 shows the effect of the span of control technology on local profitability, x_s. Increased competition lowers local profitability but more so in low-income locations. These are the locations where top firms expand and where they now compete with lower productivity firms. Although the total number of plants increases everywhere, low-income locations exhibit the largest increase in the number of plants.

Figure 10: Effect of Improvements in Span of Control Technology on λ_j and x_s



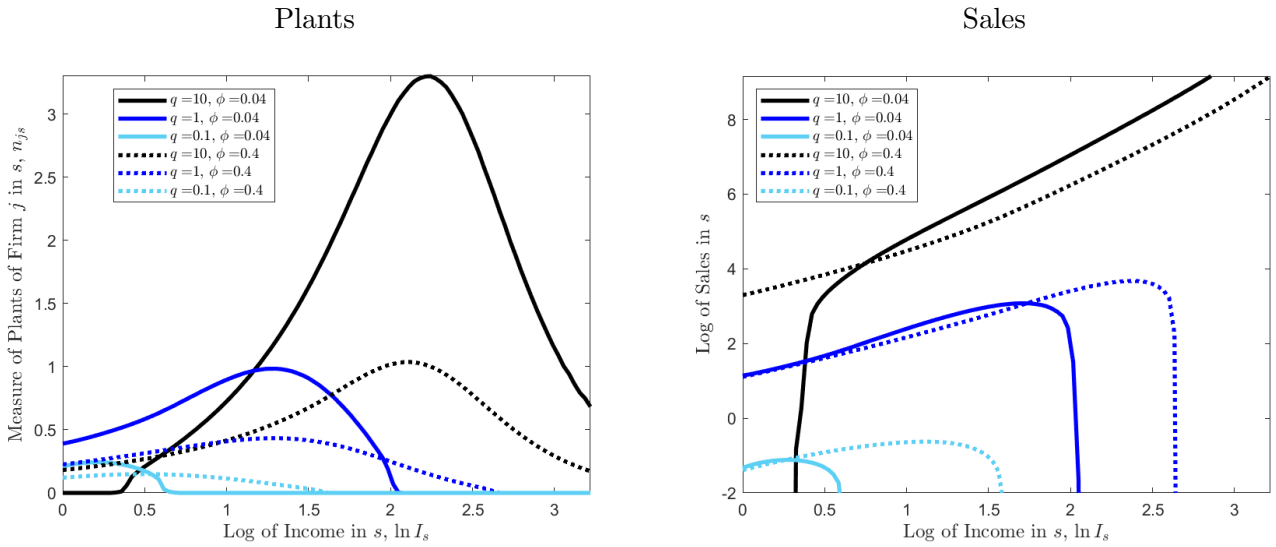
3.4.2 Improvements in an Industry's Transportation Costs

Consider the effects of an improvement in transportation technology captured by an increase in ϕ in the transportation cost function, $t(\delta) = e^{\delta/\sqrt{\phi}}$. An increase in transportation efficiency reduces the cost of reaching customers and so incentivizes firms to have fewer plants with larger catchment areas. Fewer plants imply lower managerial costs associated with firms' span of control which in turn increases productivity and induces them to expand. The larger catchment areas effectively reduce the (fixed) rent costs of serving consumers in a location, which encourages the entry of all firms in more markets but particularly incentivizes the entry of less-productive firms. Furthermore, lower transport costs imply more cannibalization between plants. This effect is particularly relevant for high-productivity firms since they operate more plants. Hence, we expect improvements in transport efficiency to disproportionately benefit low-productivity firms. The incentives to enter more locations with fewer plants implies that competition at the local level increases everywhere, as reflected by an increase in $\mathcal{Z}_s$. This countervailing force reduces firms sales in some locations.

Figure 11 shows the effect of an increase in ϕ from 0.04 to 0.4 on the mass of plants and sales of the representative high, medium and low productivity firms discussed earlier. The left-hand panel shows that all firms expand to new locations but also have fewer plants in most locations where they where already present. The top firm expands to low-income locations and now sells everywhere, while the medium and low-productivity firms expand to higher income locations. The increase in competition implies that profitability, x_s, declines almost uniformly across markets. As the right-hand side of Figure 11 shows, increased competition implies that all three firms see their sales fall in many of the markets where they were already operating.

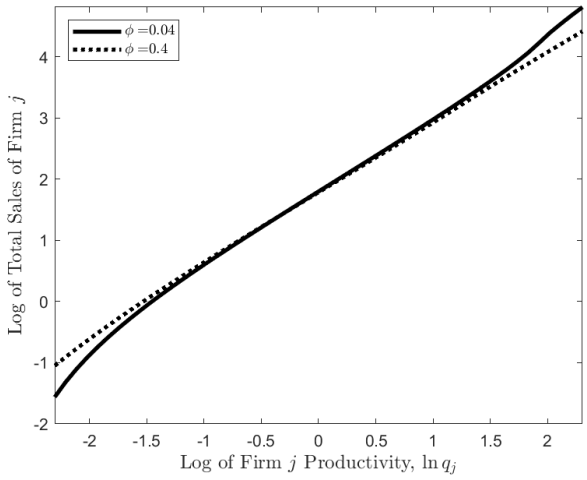
While the improvement in transport technology leads firms to expand the range of locations in which

Figure 11: Transportation Efficiency and Sorting in an Industry Equilibrium



they are active, they also have fewer plants. The effect on the total number of plants, therefore, is ambiguous. In this simulation, the total measure of plants falls in both low-income and high-income markets. However, it increases in middle-income markets as a large number of lower-productivity firms now choose to enter these markets. Overall, the improvement in transport costs favors low-productivity firms. Figure 12 shows that improvements in transportation technology lead total sales to increase for the lowest-productivity firms while total sales by the top firms decline.

Figure 12: Effect of Improvements in Transportation Efficiency on Total Firm Sales



4 Empirical Evidence

Our theory provides a number of concrete implications about the location of plants in space. In this section, we contrast these implications with U.S. evidence for the year 2014. Our main source of data is the National Establishment Time Series (NETS), which is provided by Walls & Associates. NETS provides yearly employment information for 'lines of business', which we associate to plants in the theory, and refer to as plants or establishments in the reminder of the paper.[17] For each establishment, we know its geographic coordinates, its industry classification, and its parent company.[18] We classify industries according to the SIC8 industry classification, with over 18,000 distinct industries.

We are interested in exploring how firms place their plants across space. To do so, we require a consistent definition of a 'location'. We follow Holmes and Lee (2010), and divide the continental United States into squares with side length of M miles. We present results for values of M ranging from 3 to 48 miles.

4.1 Sorting in the Data

A central and distinctive prediction of our model is that more productive firms sort towards 'better' locations. In order to study sorting in the data, we first need to map firm productivity and location characteristics to observable measures in the data. In Lemma 8, we show that firm total national employment in a given industry is strictly increasing in its productivity. We can measure a firm's total employment directly in the data. The discussion in the previous sections shows that population in a location is related to a combination of its exogenous amenities, productivity, as well as endogenous characteristics of firms that set up plants in that location. We can easily measure population density in the data (since all locations are squares with the same area), and then use this metric to rank locations.

Our main results related to sorting are presented in Proposition 4 and Proposition 6. These propositions establish that more productive firms set up relatively more plants in locations with higher land rents or higher local profitability. We do not observe local rents in the NETS data. However, using alternative data sources, it is clear that there is a very tight positive relationship between rents and our ordering of locations using population density. Figure B.1 in Section B of the Appendix shows the relationship for zip codes and counties using ACS and Zillow data. Hence, in what follows, we use population density as a measure of the local characteristics on which firms sort.

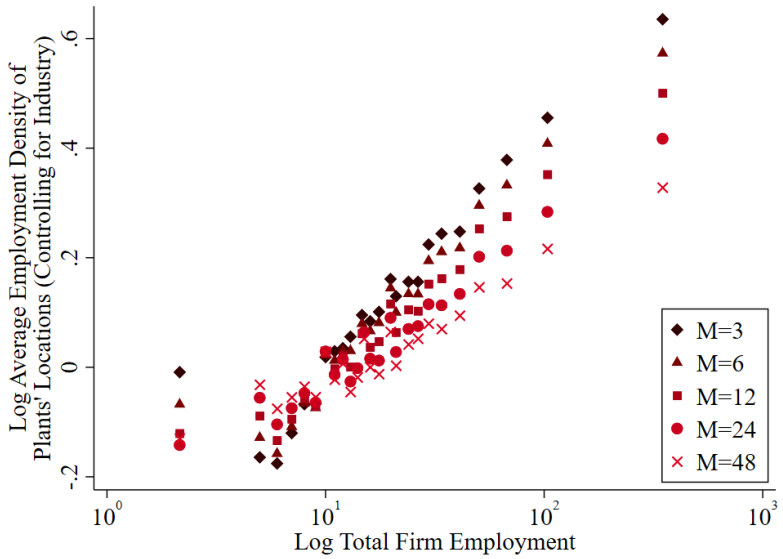
As in the theory, let $\mathcal{L}_s$, denote population density in location s. The average weighted density of the

[17] The definition of a line of business is almost identical to the definition of an establishment or plant (which we use as synonyms). An establishment may contain one or more lines of businesses. Although conceivable in principle, in practice almost all plants have a single line of business. Thus, we refer to a line of business as a plant. For those cases where two lines of business are present in the same exact location, and thus in the same plant, each line of business is identified as a single plant.

[18] A more detailed description of NETS can be found in Rossi-Hansberg et al. (2018). We mostly use only a cross-section of NETS for 2014. Compared to Census data, a cross section in NETS has an excess of very small firms, partly because it keeps track of non-employee firms. Thus, we restrict our attention to firms with at least five employees. Crane and Decker (2019) observes that NETS has imputed employment data. Once we restrict to firms with at least five employees, the fraction of plants with non-imputed employment is 81.5%. In Appendix C.5 we show that our main empirical findings regarding sorting are robust to using only the non-imputed data.

locations of a firm j operating in industry i, $\bar{L}_{ji}$, is given by $\bar{L}_{ji} \equiv \sum_s \omega_{jis} \mathcal{L}_s$, where ω_{jis} is the weight for firm j in industry i at location s. We compute the weights ω_{jis} in three alternative ways: by the number of plants owned by j in s, by its employment in s, or by assigning equal weights to each location s. For each case, the sum of the weights across locations for a given firm j operating in industry i add up to one, $\sum_s \omega_{jis} = 1$. We use weights given by the number of plants as our baseline since our results refer directly to the choice of n_{js}. Once we compute $\bar{L}_{ji}$, we subtract industry fixed effects. We use the residuals as our measurement of a firm's average weighted density of location. Figure 13 presents, for a set of different resolutions M, binned log average firm employment densities as a function of log firm size, $\ln L_j$. As the figure shows, for all resolutions, there is a strong increasing relationship between firm size and average density. As implied by Proposition 4 and Proposition 6, larger firms sort into denser places that tend to exhibit larger rents and better fundamentals.

Figure 13: Sorting in the data



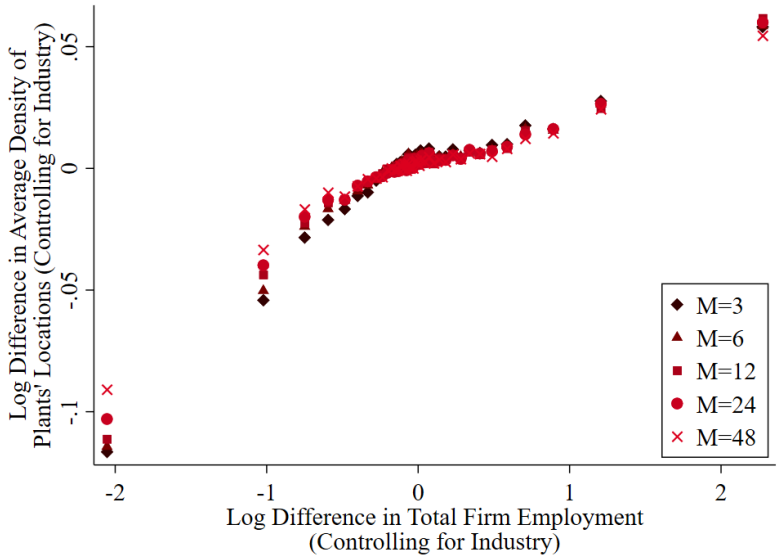
Notes: The figure presents the results for the log of the average employment density of each location, weighted by the number of establishments of a particular firm operating in a particular industry in the location. To produce the figure we first subtract industry fixed effects. We then bin the residuals by the log total firm employment at the national level. The figure presents the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

Figure C.3 in Section C of the Appendix shows that the results are robust to alternative weights. Figure C.1 in the same Appendix, implements a leave-out strategy to addresses the potential concern that the firm's presence could be driving local density. In the figure, when we compute a firm's log average density, we net out its own local employment from each local density. The resulting sorting is virtually identical.

Of course, the sorting pattern we have uncovered could arise from omitted characteristics of firms that are correlated with density. For example, if firms tend to set up plants where they are founded and denser

locations incubate more productive firms, we would obtain the pattern in Figure 13. To control for fixed firm characteristics, we can replicate the empirical sorting exercise in time changes. If a firm experiences an exogenous productivity shock that makes it grow, our theory predicts that it should shift production towards more dense locations. The productivity shock can be endogenous to all firm characteristics, including the initial distribution of plants, but we require the shock to be exogenous to the location of new plants. Using population density in 2000, we replicate the construction of the average weighted log density of the locations of a firm, but using as weights the number of plants in 2000 and in 2014 alternatively. Figure 14 plots the difference in the log average density of location for each firm between the year 2014 and the year 2000, against the log difference in the firm's total employment. It shows, at all resolutions M, that larger increases in firm size imply shifts to denser locations.[19]

Figure 14: Sorting over time



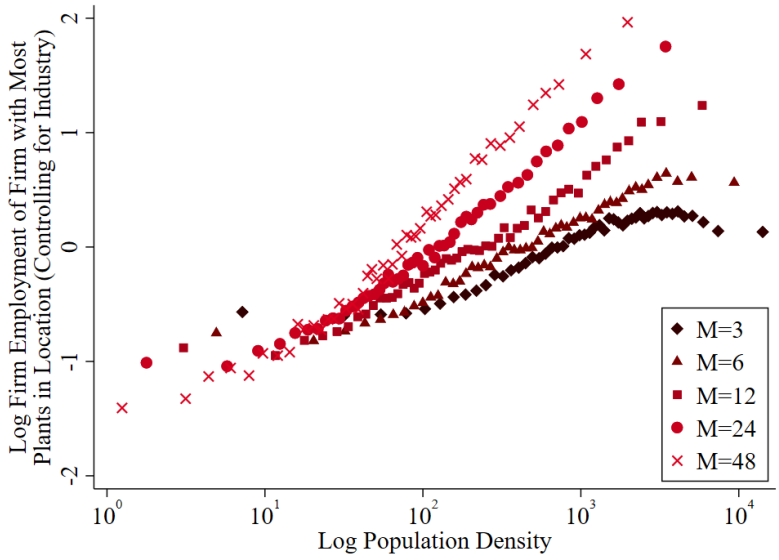
Notes: The figure is produced by calculating the average population density across all locations where a firm is present, for a given industry, holding densities at their 2000 levels and weighting by the firm's establishments in the location in both 2000 and 2014, for firms-industry pairs with positive employment in both years. We then compute the log-difference in firm size and in average population density between 2000 and 2014, regress each of them separately on industry fixed effects, and store the residuals from each regression. We bin the residuals of the log-difference in average density by the log-difference in firm size residuals. The figure presents the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

We can also explore the implications of Proposition 4 on sorting by looking at the size of the firm with the largest number of plants in each location. Sorting implies that, in locations with low population densities, low productivity firms should place more plants than large firms. Figure 15 shows that, in fact, for less

[19] These results show the *average*, across firm sizes, relationship between firm density of plant location and firm growth. Hsieh and Rossi-Hansberg (2019) show using the LBD Census data that the largest firms, in industries that are concentrating at the national level, have been progressively entering smaller cities since the late 70's. Both results are consistent since top firms in these industries are behaving differently than the average firm. Proposition 4 admits this case, if the rent R^* is higher than any existing rent when we compare the location of the plants of the top firm before and after its change in productivity.

dense locations, the firm with most local plants has a smaller number of total employees. Specifically, the national size of the firm with the most plants in a location increases with population density.[20] In Figure 15, there are locations where multiple firms tie for the highest number of plants. In those cases, we use the average national firm size among these firms. In Section C in the Appendix, we show that the finding is robust to dropping cases with ties or to using the national size of the largest firm among those tied.

Figure 15: The national size of the largest firm in town



Notes: The figure is produced by finding the log employment of the firm with the most plants in an industry and location, and regressing its total employment on industry fixed effects, weighted by each industry's total employment. In locations where multiple firms are tied for the highest number of plants, we take the average of the firm size. We then bin the residuals against log population density for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

In Section C.4 of the Appendix, we show that these sorting patterns are also present when we focus on particular industry aggregates. Specifically we reproduce the results using data for Manufacturing, Services, or Retail Trade exclusively. We find that the strength of sorting varies across sectors. It is weak in Manufacturing, moderate for Services, and strong for Retail Trade. This ranking is expected since transport costs in manufacturing are much lower. We revisit this implication later in this section. Finally, Appendix C.5 shows that our results are robust to using only non-imputed data.

4.2 The Role of Span of Control Costs

In the multi-plant firm problem we analyzed in Section 2, two firms present in the same location can choose to have a different number of plants either because they have different productivity, or because they have a different cost of increasing their span of control, λ_j. Lemma 3 shows that higher productivity firms have

[20] Note that the relationship is clearly steeper as we increase the resolution M. This is simply the mechanical result of spatial averaging since larger M implies a lower range of densities but the range of firm employment sizes remains the same.

a higher cost of increasing their span of control, but the marginal cost of increasing the number of plants, λ_j, depends on the entire distribution of local characteristics. Now consider two firms that have the same number of plants in a location. The larger cost of setting up additional plants, in any location, faced by more productive firms implies that plants operated by these firms will be larger. Formally, we can write the average plants size of firm j in location s as

$$\bar{l}_{js} = (\varepsilon - 1) z_j^{\varepsilon - 1} \frac{x_s}{W_s} \frac{\kappa\left(n_{js}\right)}{n_{js}}. \tag{8}$$

Then, it is straightforward to see that, if $z_j > z_{\tilde{j}}$, then $\bar{l}_{js} > \bar{l}_{\tilde{j}s}$ in locations where $n_{js} = n_{\tilde{j}s}$. Hence, controlling for the measure of plants in a location, more productive firm place larger plants in a given location. Consistent with this logic, Figure 16 presents the positive relationship between average plant employment and total firm employment, after controlling for the number of plants and industry-location fixed effects.[21]

4.3 Local Saturation

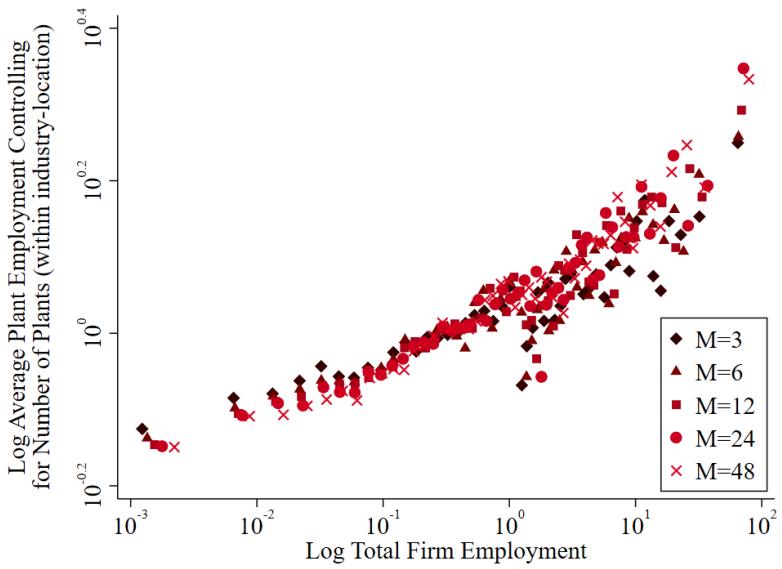
As evident from equation (8), a firm's average plant size in a location s, $\bar{l}_{js}$, depends on its productivity, z_j, local characteristics, x_s/W_s, and the number of plants, n_{js} through the local efficiency of distribution function, κ. Lemma 2 describes the properties of the efficiency of distribution. It starts at 0, where it increases linearly because cannibalization is limited with few plants, becomes increasing and concave as the number of plants grows, and eventually converges to 1. The function describes how the firm increasingly saturates local markets as it adds plants. Taking logs and the total derivative of equation (8), we obtain that

$$d \ln \bar{l}_{js} = d \ln z_j^{\varepsilon - 1} + d \ln \frac{x_s}{W_s} + \left(\frac{n_{js} \kappa'(n_{js})}{\kappa(n_{js})} - 1 \right) d \ln n_{js} \ , \tag{9}$$

where $\lim_{n \to 0} \frac{n_{js}\kappa'(n_{js})}{\kappa(n_{js})} = 1$ and $\lim_{n \to \infty} \frac{n_{js}\kappa'(n_{js})}{\kappa(n_{js})} = 0$ by Lemma 2. The equation implies that, controlling for changes in firm productivity and local characteristics, firms with very few local plants that add plants in a location make few adjustments to the size of their plants; with no saturation, new plants do not affect the size of plants that already exist. In contrast, firms that already own many local plants significantly reduce the size of their plants when adding plants to a location. In the limit, as the market becomes completely saturated, any increase in the number of plants is compensated by a similar proportional reduction in plant size. Hence, at least on average over the whole range, the correlation between changes in plant size and changes in the number of plants is decreasing in the number of plants, conditional on appropriate firm and

[21] These results are consistent with those found in Fernandes et al. (2018), which finds that a large fraction of the variation of exports in bilateral trade is through the intensive margin of trade. In our model, this variation maps into variation in average plant employment within a location. Moreover, our empirical findings are inconsistent with the application of trade models relying on Pareto distributions to explain the way firms locate their plants across space within the US (e.g., Lind and Ramondo, 2018).

Figure 16: Span of Control



Notes: The figure presents the results for the log of the average plant employment of a firm within a location. To produce the figure we first subtract industry-location fixed effects, and we control for the number of plants of the firm in the location. We then bin the residuals by the log total firm employment at the national level, which is controlled for industry fixed effects. Also, we subtract the own firm contribution of employment in a location from that firm's total employment. The figure presents the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

location controls. To test this prediction, we estimate,

$$\Delta \ln \bar{l}_{js} = \beta_0 \ln n_{js,2000} + \beta_1 \Delta \ln n_{js} + \beta_2 \ln n_{js,2000} \times \Delta \ln n_{js} + FE_s + FE_j + \varepsilon_{js} \ ,$$

where FE_s and FE_j account for trends in location and firm characteristics, $\ln n_{js,2000}$ denotes the log of j's number of plants in location s in year 2000, and $\Delta \ln n_{js}$ is the log change in the number of plants between 2000 and 2014. The theoretical implication is that β_2 is negative. Table I presents the results. As the table shows, $\beta_2 < 0$ for all resolutions M. The effect of the number of plants on plant employment size is more negative the more plants a firm has in that location.

4.4 The Role of Local Characteristics

The decision of how many plants to set up in a particular location depends, of course, on the characteristics of the area. The first order condition (3) identifies the level of local profitability, x_s, and local rents, R_s, as the relevant local variables. Proposition 6 tells us that more productive firms sort into locations with higher local profitability, and we provided evidence for this claim in this section already. We are now interested in understanding how changes in local profitability, that lead to a change in rents, affect the number of plants

Table I: Local Saturation and Firm Growth

	(1) $\Delta \ln \bar{l}_{js}$	(2) $\Delta \ln \bar{l}_{js}$	(3) $\Delta \ln \bar{l}_{js}$	(4) $\Delta \ln \bar{l}_{js}$	(5) $\Delta \ln \bar{l}_{js}$
$\ln n_{js,2000}$	-0.0792***	-0.0467***	-0.0402***	-0.0634***	-0.0661***
	(0.0260)	(0.0169)	(0.0136)	(0.0115)	(0.0101)
$\Delta \ln n_{js}$	0.0729***	0.0460**	0.0768***	0.0610***	0.0639***
	(0.0262)	(0.0190)	(0.0157)	(0.0135)	(0.0126)
$\ln n_{js,2000} \times \Delta \ln n_{js}$	-0.0954***	-0.00369	-0.0194	-0.0192**	-0.0225***
	(0.0308)	(0.0203)	(0.0132)	(0.00970)	(0.00787)
Observations	20,230	30,583	41,246	49,888	56,170
R-squared	0.628	0.621	0.604	0.589	0.571
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
SIC8-firm FE	Yes	Yes	Yes	Yes	Yes
M	3	6	12	24	48

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table presents the results of regressing the change from 2000 to 2014 in the log average plant employment of a firm in a location on the initial level of the log of the number of plants of the firm in 2000, the change from 2000 to 2014 in the log number of plants of the firm in the location, and the interaction of the initial level of the log of the number of plants of the firm in the location and the change from 2000 to 2014 in the log number of plants of the firm in the location. The regression also includes industry-location and industry-firm fixed effects. The table presents the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

of firms with different productivity levels.

To explore the implications of optimal firm behavior, differentiate the log of condition (3) to obtain[22]

$$\frac{d \ln n_{js}}{d \ln x_s} = -\frac{\kappa'(n_{js})}{n_{js}\kappa''(n_{js})}\left[1 - \frac{R_s}{R_s + \lambda_j}\frac{d \ln R_s}{d \ln x_s}\right] > 0 \ .$$

We know from Lemma 3 that more productive firms have higher span of control costs, λ_j. Furthermore, note that, if $d \ln R_s / d \ln x_s > 0$, the term in brackets is increasing in λ_j. Hence, conditional on the number of plants, since κ is increasing and concave, more productive firms set up and operate relatively more plants as local profitability increases. The intuition is, as before, that rents make up a smaller portion of the cost of setting new plants for the more productive firms (since their larger span of control also implies a larger λ).

We explore this implication empirically in Table II where we regress the growth in the number of establishments of a firm in a location s, on the log of its number of establishments in s in 2000, $\ln n_{js,2000}$, an interaction of the log of total firm size (our proxy for firm profitability) and the change in local log density (our proxy for the change in local characteristics), $\ln L_{j,2000} \times \Delta \ln \mathcal{L}_s$, as well as firm-industry and location-industry fixed effects.[23] The results confirm that, as implied by the theory, the interaction is positively

[22] The location is small in the firm's problem and so changes in x_s do not affect λ_j.

[23] We compute the growth rate of the number of plants n_{js} as in (see, Davis et al., 1996), namely, we let Growth of $n_{js} = (n_{js,2014} - n_{js,2000})/(0.5n_{js,2014} + 0.5n_{js,2000})$. This growth measure is particularly useful when entry and exit of firms, and therefore all their plants, across locations is substantial. This approach allows us to minimize the loss of information that would

related with growth in the number of plants. Namely, conditional on the number of plants, larger firms at the national level add more plants when local conditions improve.

Table II: Rents and Firm Growth

	(1) Growth in n_{js}	(2) Growth in n_{js}	(3) Growth in n_{js}	(4) Growth in n_{js}	(5) Growth in n_{js}
$\ln n_{js,2000}$	0.00438	-0.00302	-0.00213	0.00150	0.00511*
	(0.00646)	(0.00457)	(0.00357)	(0.00303)	(0.00268)
$\ln L_{j,2000} \times \Delta \ln \mathcal{L}_s$	0.0100*	0.0126**	0.0171***	0.0237***	0.0427***
	(0.00540)	(0.00522)	(0.00549)	(0.00622)	(0.00717)
Observations	272,506	360,721	418,772	443,227	442,352
R-squared	0.793	0.788	0.782	0.780	0.775
SIC8-firm FE	Yes	Yes	Yes	Yes	Yes
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
M	3	6	12	24	48

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table presents the results of regressing the growth from 2000 to 2014 in the number of plants of a firm in a location on the log of the number of plants of the firm in a location in 2000, and on the interaction of the log of the firm's employment at the national level in 2000 and the change from 2000 to 2014 in the log of the location's population density. The regression also includes industry-location and industry-firm fixed effects. We compute the growth in the number of plants of a firm in a location as Growth of $n_{js} = (n_{js,2014} - n_{js,2000})/(0.5n_{js,2014} + 0.5n_{js,2000})$ (see, Davis et al., 1996). The table presents the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

4.5 The Role of Transportation Efficiency

We are interested in understanding how the transportation infrastructure of a region affects the location of plants, as well as their size. To that end, we parameterize transport costs using a measure of transport efficiency, ϕ_{js}. Specifically, we let the cost of delivering a product at distance δ, when local transport efficiency is ϕ, be given by $t(\delta;\phi) \equiv t\left(\frac{\delta}{\sqrt{\phi}}\right)$, as in Section 3.4. Thus, as before, the corresponding local efficiency of distribution function becomes $\kappa(n;\phi) \equiv \kappa(\phi n)$ and the first order condition in condition (3) becomes

$$z_j^{\varepsilon-1} x_s \phi_{js} \kappa'(\phi_{js} n_{js}) = R_s + \lambda_j \ . \tag{10}$$

Hence, holding fixed z_j, x_s, R_s, and λ_j, we obtain that

$$\frac{d \ln n_{js}}{d \ln \phi_{js}} = -1 - \frac{\kappa'(\phi_{js} n_{js})}{\phi_{js} n_{js} \kappa''(\phi_{js} n_{js})}. \tag{11}$$

There are two competing forces that influence how an improvement in transport efficiency alters firm j's choice of n_{js}. First, given total sales, the firm moves towards larger plants to supply its customers using

follow from using more standard growth measures.

larger catchment areas. This force is associated with a proportional reduction in the number of plants reflected in the term -1. Second, since transport is now cheaper, the plant can reach consumers at lower cost which incetivizes it to raise its sales and employment. The firm, therefore, sets up a greater number of production plants, captured by the second term which, given the concavity of κ, is positive. The resulting net effect is generally ambiguous. However, the properties of κ in Lemma 2 imply that this effect is decreasing in n_{js}, and therefore decreasing in total local employment (i.e. equation (8)), firm productivity, and total firm size. More precisely, we can show that $\lim_{n\to 0} -1 - \frac{\kappa'(n)}{n\kappa''(n)} = \infty$ and $\lim_{n\to\infty} -1 - \frac{\kappa'(n)}{n\kappa''(n)} = -\frac{1}{3}$. In sum, better transport efficiency leads to a more pronounced reduction in the number of plants (or a less pronounced increase) for firms with larger local footprints.

Consider also the effect of transport efficiency on average plant size. Using equations (8) and (10), we obtain that

$$\frac{d\ln \bar{l}_{js}}{d\ln \phi_{js}} = \underbrace{\frac{-\kappa'\left(n_{js}\phi_{js}\right)^2}{\kappa\left(n_{js}\phi_{js}\right)\kappa''\left(n_{js}\phi_{js}\right)}}_{\frac{d\ln L_{js}}{d\ln \phi_{js}}} - \underbrace{\left(-1 - \frac{\kappa'\left(n_{js}\phi_{js}\right)}{\phi_{js}n_{js}\kappa''\left(n_{js}\phi_{js}\right)}\right)}_{\frac{d\ln n_{js}}{d\ln \phi_{js}}} . \tag{12}$$

The first term is the direct effect of sales to consumers holding fixed the mass of plants; employment increases so as to raise production to satisfy each consumer's increased demand. This effect is always positive. In addition, as discussed above, average employment may also change as the measure of plants adjusts following better transport efficiency. The properties of κ in Lemma 2 imply that the first term is large when n_{js} is small and small in the limit as n_{js} becomes large. In contrast, the discussion above implies that $-\frac{d\ln n_{js}}{d\ln \phi_{js}}$ is increasing in n_{js}. Large firms with many plants expand total local employment by less than smaller firms following an improvement in transport efficiency, but they are also less aggressive in setting up new plants. The final effect of firm size on average plant size is, therefore, ambiguous.

We now go to the data to determine whether, in fact, the effect of transportation efficiency on the number of plants is decreasing in firms' local employment. Without directly identifying exogenous changes in ϕ_{js}, we can nevertheless assess the effect of transportation efficiency in the cross-section, controlling for firm and location fixed effects. Specifically, we estimate

$$\ln n_{js} = \gamma_0 \ln L_{js} + \gamma_1 \ln L_{js} \times \ln \phi_{js} + FE_s + FE_j + \varepsilon_{js} , \tag{13}$$

where FE_s and FE_j again denote location-industry and firm-industry fixed effects. The prediction discussed above implies that $\gamma_1 < 0$.[24]

To estimate equation (13), we require a local, firm-specific, measure of ϕ_{js}. We can also exploit variation across industries, since firms operate in different industries. We propose seven distinct measures of transportation efficiency, described in detail in Table IV in Section D of the Appendix. The first two, the Gini

[24]The prediction of our theory applies holding x_s, R_s, z_j, and λ_j constant. This regression controls for firm and location fixed effects, but uses variation in transportation efficiency that, we would expect, affects these firm and local characteristics in general equilibrium. Hence, our interpretation of the results relies on the changes in these firm or local variables being sufficiently small.

coefficient of an industry's employment across space, and the Ellison and Glaeser (1997) index (which corrects the Gini coefficient for granularity of individual plants), measure industry concentration across space. More concentration is likely the result of smaller transport costs or higher transportation efficiency. Admittedly, these measures are arguably endogenous to the number of plants. The third measure, (Agarwal et al., 2017) consumer gravity, uses credit card purchases to estimate, for each industry, the elasticity of purchases with respect to distance from home; we interpret a lower elasticity as indicating low transport costs. Our fourth measure, uses measures of average freight costs computed by Bernard et al. (2006); while this measure is closest to a primitive measure of trade costs, it is only available for manufacturing industries. Fifth, we use measures of trade costs from Gervais and Jensen (2019) for service and manufacturing industries inferred from local disparities in supply and demand for each industry. Sixth, we use a measure of traveling speed computed by Couture et al. (2018) for 50 MSAs for 2008. Seventh, we use a measure of traveling time for 470 MSAs, computed by the Texas Transportation Institute. Each measure is standardized and built such that ϕ_{js} is increasing in each measure.

Table III: The Effect Transportation Efficiency on Number of Plants in a Location

VARIABLES	(1) $\ln n_{js}$	(2) $\ln n_{js}$	(3) $\ln n_{js}$	(4) $\ln n_{js}$	(5) $\ln n_{js}$	(6) $\ln n_{js}$	(7) $\ln n_{js}$	(8) $\ln n_{js}$
$\ln L_{js}$	0.276***	0.244***	0.276***	0.351***	0.129***	0.271***	0.271***	0.272***
	(0.00145)	(0.00131)	(0.00145)	(0.00219)	(0.00555)	(0.00152)	(0.00183)	(0.00150)
X Gini		-0.151***						
		(0.00171)						
X Ellison-Glaeser			-0.0413***					
			(0.00607)					
X Consumer Gravity				-0.0970***				
				(0.00150)				
X Freight Cost					-0.0200***			
					(0.00584)			
X Trade Cost						-0.0250***		
						(0.00118)		
X Speed Score							-0.00355***	
							(0.00131)	
X Travel Time								-0.00881***
								(0.00117)
Observations	366,979	366,979	366,979	209,700	8,166	345,771	207,155	323,782
R-squared	0.747	0.769	0.748	0.798	0.692	0.750	0.769	0.753
SIC8-location FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC8-firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
M	24	24	24	24	24	24	24	24

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table presents the regression results of (13), including location and firm fixed effects. The transportation efficiency measures that we use are described in Table IV. All transportation efficiency measures are standardized. For this table, we use resolution $M = 24$.

Table III presents the results of running the cross-sectional regression presented in (13) for alternative transportation efficiency measures and $M = 24$. While the measures capture different aspects of transporta-

tion efficiency (shipping costs, distribution costs, and consumer travel costs), the regression results are all consistent. As implied by the theory, better transportation efficiency has a larger negative impact on the number of plants of firms with more local employment.[25]

5 Conclusions

In this paper, we propose a novel methodology to analyze the problem of how to serve customers distributed across heterogeneous locations when firms face transport costs, fixed costs of setting up new plants, and span of control costs of managing multiple plants. Although the basic trade-off between transport costs and cannibalization is clear, characterizing the solution to this core problem in economics has proven elusive given its complexity. In order to make progress, we propose a limit problem in which firms choose a density of plants in space. A large combinatorial problem is therefore reduced to a much simple calculus of variations problem. The solution to this problem can be easily characterized and the problem can be readily incorporated in a general equilibrium spatial setup with labor mobility, as we have demonstrated.

The solution to the firm problem has a number of unique predictions. First, and most important, is that firms sort in space. Specifically, more productive firms operate relatively more plants in locations with better characteristics or simply higher rents. Less productive firms, in turn, operate relatively more plants in worse, low rent, locations. Furthermore, conditional on the number of plants, more productive firms operate larger plants. These and other predictions of the theory are empirically verified using NETS establishment level data for 2000 to 2014 in the U.S.

The methodology proposed in this paper can readily be used to understand the role of changes in transport infrastructure on plant locations. We illustrate numerically how firms in a 'small' industry – one that does not affect local rents or wages – adjust by opening fewer plants but in more locations. In the data, we find that local transport infrastructure improvements indeed lead to a lower number of plants, particularly for the most productive firms. We carry out a similar quantitative exercise to illustrate the effects of improvements in the span of control technology, where we see large firms expanding into low-rent markets. Studying general equilibrium counterfactuals for 'large' industries that affect local factor prices, or for the whole economy, is left for future research. A quantitative general equilibrium analysis of such changes could be used to study the implications of secular technological changes for the spatial distribution of economic activity, as well as local competition and concentration. These are exciting avenues that our methodology now makes feasible.

[25]The results are very stable across resolutions, as shown in Figure D.1 in Section D of the Appendix, where we present a plot of the coefficients on the interaction term, γ_1.

References

Agarwal, S., Jensen, J. B., and Monte, F. (2017). The Geography of Consumption. NBER Working Papers 23616, National Bureau of Economic Research, Inc.

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J., and Li, H. (2019). A Theory of Falling Growth and Rising Rents. Working Paper Series 2019-11, Federal Reserve Bank of San Francisco.

Arkolakis, C. and Eckert, F. (2017). Combinatorial discrete choice. Technical report.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553.

Bernard, A. B., Jensen, J. B., and Schott, P. K. (2006). Trade costs, firms and productivity. *Journal of Monetary Economics*, 53(5):917–937.

Bilal, A. and Rossi-Hansberg, E. (2019). Location as an asset. Technical report.

Bollobas, B. (1973). The optimal arrangement of producers. *Journal of the London Mathematical Society*.

Cao, D., Sager, E., Hyatt, H., and Mukoyama, T. (2019). Firm Growth through New Establishments. 2019 Meeting Papers 1484, Society for Economic Dynamics.

Christaller, W. (1933). *Die zentralen Orte in Suddeutschland*.

Couture, V., Duranton, G., and Turner, M. A. (2018). Speed. *The Review of Economics and Statistics*, 100(4):725–739.

Crane, L. and Decker, R. (2019). Business Dynamics in the National Establishment Time Series (NETS)/Leland Crane, Ryan Decker. Finance and economics discussion series, Board of Governors of the Federal Reserve System (US).

Davis, D. R. and Dingel, J. I. (2019). A spatial knowledge economy. *American Economic Review*, 109(1):153–70.

Davis, S. J., Haltiwanger, J., and Schuh, S. (1996). Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts. *Small Business Economics*, 8(4):297–315.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, 106(3):479–524.

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

Eeckhout, J., Pinheiro, R., and Schmidheiny, K. (2014). Spatial sorting. *Journal of Political Economy*, 122(3):554–620.

Ellison, G. and Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105(5):889–927.

Fejes Toth, L. (1953). *Lagerungen in der Ebene, auf der Kugel und im Raum*. Springer-Verlag.

Fernandes, A. M., Klenow, P. J., Meleshchuk, S., Pierola, D., and Rodrguez-Clare, A. (2018). The Intensive Margin in Trade. NBER Working Papers 25195, National Bureau of Economic Research, Inc.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11):3117–53.

Gervais, A. and Jensen, J. B. (2019). The tradability of services: Geographic concentration and trade costs. *Journal of International Economics*, 118(C):331–350.

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

Holmes, T. J. and Lee, S. (2010). Cities as Six-by-Six-Mile Squares: Zipf's Law? In *Agglomeration Economics*, NBER Chapters. National Bureau of Economic Research, Inc.

Hopenhayn, H. A. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, 60(5):1127–1150.

Hsieh, C.-T. and Rossi-Hansberg, E. (2019). The industrial revolution in services. Working Paper 25968, National Bureau of Economic Research.

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

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

Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica*, 50(3):649–670.

Lind, N. and Ramondo, N. (2018). Trade with Correlation. NBER Working Papers 24380, National Bureau of Economic Research, Inc.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Ramondo, N. (2014). A quantitative approach to multinational production. *Journal of International Economics*, 93(1):108 – 122.

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

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9(1):21–58.

Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2018). Diverging Trends in National and Local Concentration. NBER Working Papers 25066, National Bureau of Economic Research, Inc.

Rossi-Hansberg, E. and Wright, M. L. J. (2007). Establishment size dynamics in the aggregate economy. *American Economic Review*, 97(5):1639–1666.

Tintelnot, F. (2016). Global production with export platforms. *The Quarterly Journal of Economics*, 132(1):157–209.

Ziv, O. (2019). The micro-geography of productivity, density, and sorting. Technical report.

Appendix

A Proofs

A.1 Properties of the Efficiency of Distribution

In this appendix, we provide a proof of Lemma 2. First, define $\tilde{t}(\delta) \equiv t(\delta)^{1-\varepsilon}$. A hexagon with area x has sides of length $l = \psi\sqrt{x}$, where $\psi \equiv 2^{1/2}3^{-3/4}$. The integral of the function $\tilde{t}\left(\|s\|\right)$ over a hexagon with area x can be expressed as

$$g(x) = \int_0^{\psi x^{1/2}} \varpi\left(\frac{\delta}{\psi x^{1/2}}\right)\tilde{t}(\delta)2\pi\delta d\delta$$

where $\varpi(r)$ is the fraction of circle with radius r that intersects with a hexagon with side length 1. That is, if $\alpha \equiv \sqrt{3}/2$ is the radius of the largest circle that can be inscribed in a hexagon with side length 1, then $\varpi(r) = 1$ for $r \in [0, \alpha]$, $\varpi'(r) < 0$ for $r \in (\alpha, 1)$, and $\varpi(1) = 0$.[26] We first rewrite g in a form that is easier to manipulate. First, define $\tilde{t}(\delta) \equiv t(\delta)^{1-\varepsilon}$. We then can change variables

$$\begin{aligned} g(x) &= \int_0^{\psi\sqrt{x}} \varpi\left(\frac{\delta}{\psi\sqrt{x}}\right)\tilde{t}(\delta)2\pi\delta d\delta \\ &= \psi^2 x\int_0^1 \varpi(u)\tilde{t}\left(\psi\sqrt{x}u\right)2\pi u du \end{aligned}$$

This implies that

$$\kappa(n) = ng\left(\frac{1}{n}\right) = n\int_0^{\psi n^{-1/2}} \varpi\left(\frac{\delta}{\psi n^{-1/2}}\right)\tilde{t}(\delta)2\pi\delta d\delta \tag{14}$$

$$\kappa(n) = ng\left(\frac{1}{n}\right) = \psi^2\int_0^1 \varpi(u)\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du \tag{15}$$

It will be useful to have expressions for the first and second derivative Differentiating with respect to n yields

$$\kappa'(n) = \psi^2\int_0^1 \varpi(u)\tilde{t}'\left(\psi n^{-1/2}u\right)\left(-\psi\frac{1}{2}n^{-3/2}u\right)2\pi u du \tag{16}$$

[26] What is ϖ? To get at this, for a hexagon with side length 1, a circle with radius $\delta = \sqrt{1-(1/2)^2} = \frac{\sqrt{3}}{2}$ will be full inscribed. Consider a circle with radius between $\delta \in \left(\frac{\sqrt{3}}{2}, 1\right)$. What fraction of the circle is inside the hexagon? Consider two line segments, each emanating from the center of the hexagon to the border of the hexagon. One of length $\frac{\sqrt{3}}{2}$ which is perpendicular to the side of the hexagon, and one of length δ. The angle θ between the two satisfies $\cos(\theta) = \frac{\sqrt{3}/2}{\delta}$. The fraction of the circle of length δ that is outside the hexagon is therefore $\frac{12\theta}{2\pi}$. Therefore $\varpi(\delta) = \begin{cases} 1 & 0 \leq \delta \leq \sqrt{3}/2 \\ 1 - \frac{6}{\pi}\cos^{-1}\left(\frac{\sqrt{3}/2}{\delta}\right) & \sqrt{3}/2 \leq \delta \leq 1 \end{cases}$.

To find the second derivative, we change variables once more to get

$$\kappa'(n) = \int_0^{\psi n^{-1/2}} \varpi\left(\frac{\delta}{\psi n^{-1/2}}\right)\left[-\tilde{t}'(\delta)\right]\pi\delta^2 d\delta$$

Differentiating once more, using $\varpi(1) = 0$, and changing variables yields

$$\begin{aligned}\kappa''(n) &= \int_0^{\psi n^{-1/2}} \varpi'\left(\frac{\delta}{\psi n^{-1/2}}\right)\frac{\delta}{\psi}\frac{1}{2}n^{-\frac{1}{2}}\left[-\tilde{t}'(\delta)\right]\pi\delta^2 d\delta \\ &= \psi^3 n^{-5/2}\frac{\pi}{2}\int_0^1 \varpi'(u)u^3\left[-\tilde{t}'\left(\psi n^{-1/2}u\right)\right]du\end{aligned}$$

Using the fact that $\varpi'(r) = 0$ for $r \in (0, \alpha)$ gives

$$\kappa''(n) = \psi^3 n^{-5/2}\frac{\pi}{2}\int_\alpha^1 \varpi'(u)u^3\left[-\tilde{t}'\left(\psi n^{-1/2}u\right)\right]du \tag{17}$$

Claim A.1 *$\kappa(n) \equiv ng\left(\frac{1}{n}\right)$ is strictly increasing and strictly concave, and satisfies the following properties:*

1. *$\kappa(0) = 0$;*
2. *$\lim_{n\to\infty}\kappa(n) = 1$;*
3. *$1-\kappa(n)$ follows a power law with exponent $\frac{1}{2}$ as $n \to \infty$, i.e., $\lim_{n\to\infty}\sqrt{n}\left[1-\kappa\left(n\right)\right] = -\tilde{t}'(0)\frac{\sqrt{2}}{3^{3/4}}\left(\frac{1}{3}+\frac{\ln 3}{4}\right) > 0$.*

Proof. (16) implies that κ' is strictly positive because $t' > 0$ and $\varepsilon > 1$ imply that $\tilde{t}' < 0$. (17) implies that κ'' is strictly negative because ϖ' is strictly negative on $(\alpha, 1)$. $\kappa(0) = 0$ follows from (15) and the fact that $\lim_{y\to\infty} t(y) = \infty$ which implies that $\lim_{y\to\infty}\tilde{t}(y) = 0$. $\lim_{n\to\infty}\kappa(n) = 1$ follows from (15) and the facts that $\tilde{t}(0) = 1$, and $\psi^2\int_0^1\varpi(u)2\pi u du = 1$.

Beginning with (15), we can express $\sqrt{n}[1-\kappa(n)]$ as $\sqrt{n}\left(1-\psi^2 2\pi\int_0^1\varpi(u)\tilde{t}\left(\frac{\psi u}{\sqrt{n}}\right)udu\right)$. Taking the limit as $n\to\infty$, using $x = \sqrt{n}$, and using L'Hopital's rule gives

$$\begin{aligned}\lim_{n\to\infty}\sqrt{n}\left(1-\kappa(n)\right) &= \lim_{n\to\infty}\sqrt{n}\left(1-\psi^2 2\pi\int_0^1\varpi(u)\tilde{t}\left(\frac{\psi u}{\sqrt{n}}\right)udu\right) \\ &= \lim_{x\to 0}\frac{1-\psi^2 2\pi\int_0^1\varpi(u)\tilde{t}\left(\psi u x\right)udu}{x} \\ &= \lim_{x\to 0}\frac{-\psi^2 2\pi\int_0^1\varpi(u)\tilde{t}'\left(\psi u x\right)\psi u^2 du}{1} \\ &= \left[-\tilde{t}'(0)\right]\psi^2 2\pi\int_0^1\varpi(u)\psi u^2 du\end{aligned}$$

The result follows from the fact that $\psi^2 2\pi\int_0^1\varpi(u)\psi u^2 du = \psi\left(\frac{1}{3}+\frac{\ln 3}{4}\right) = 2^{1/2}3^{-3/4}\left(\frac{1}{3}+\frac{\ln 3}{4}\right)$ ■

Before proceeding, it will be useful to derive an alternative expression for $\kappa'(n)$. Differentiating (14) with respect to n yields

$$\begin{aligned}\kappa'(n) &= \int_0^{\psi n^{-1/2}} \varpi\left(\frac{\delta}{\psi n^{-1/2}}\right)\tilde{t}(\delta)2\pi\delta d\delta \\ &\quad +n\varpi(1)\tilde{t}\left(\psi n^{-1/2}\right)2\pi\psi n^{-1/2}\left(-\frac{1}{2}\right)\psi n^{-3/2} \\ &\quad +n\int_0^{\psi n^{-1/2}} \varpi'\left(\frac{\delta}{\psi n^{-1/2}}\right)\frac{\delta}{\psi}\frac{1}{2}n^{-1/2}\tilde{t}(\delta)2\pi\delta d\delta\end{aligned}$$

Noting that $\varpi(1)=0$ and changing variables gives

$$\begin{aligned}\kappa'(n) &= \int_0^{\psi n^{-1/2}}\left[\varpi\left(\frac{\delta}{\psi n^{-1/2}}\right)+\frac{1}{2}\varpi'\left(\frac{\delta}{\psi n^{-1/2}}\right)\frac{\delta}{\psi n^{-1/2}}\right]\tilde{t}(\delta)2\pi\delta d\delta \\ &= \frac{\psi^2}{n}\int_0^1\left[\varpi(u)+\frac{1}{2}\varpi'(u)u\right]\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du\end{aligned}$$

We can separate this into two terms, the integral over $u\in[0,\alpha]$ and the integral from $[\alpha,1]$. For $u\in[0,\alpha)$, $\varpi(u)=1$ and $\varpi'(u)=0$, so we can express the integral as

$$\kappa'(n)=\frac{\psi^2}{n}\int_0^\alpha\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du+\frac{\psi^2}{n}\int_\alpha^1\left[\varpi(u)+\frac{1}{2}\varpi'(u)u\right]\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du \tag{18}$$

Claim A.2 *If* $\lim_{\delta\to\infty}\tilde{t}(\delta)\delta^2=0$ *then* $\kappa'(0)=\int_0^\infty\tilde{t}(\delta)2\pi\delta d\delta$

Proof. Taking the limit of the first term of (18) gives

$$\lim_{n\to\infty}\frac{\psi^2}{n}\int_0^\alpha\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du=\lim_{n\to\infty}\int_0^{\alpha\psi n^{-1/2}}\tilde{t}(\delta)2\pi\delta d\delta=\int_0^\infty\tilde{t}(\delta)2\pi\delta d\delta$$

The second term of (18) can be expressed as

$$\frac{\psi^2}{n}\int_\alpha^1\left[\varpi(u)+\frac{1}{2}\varpi'(u)u\right]\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du=\int_\alpha^1\left[\frac{\varpi(u)}{u}+\frac{1}{2}\varpi'(u)\right]\tilde{t}\left(\psi n^{-1/2}u\right)\left(\psi n^{-1/2}u\right)^2 2\pi du$$

We next show that the limit of this second term is zero. If $\lim_{x\to\infty}\tilde{t}(x)x^2=0$, then $\tilde{t}(x)x^2$ has a peak. call it $\bar{r}$. Then the function $\left[\frac{\varpi(u)}{u}+\frac{1}{2}\varpi'(u)\right]\tilde{t}\left(\psi n^{-1/2}u\right)\left(\psi n^{-1/2}u\right)^2 2\pi$ is dominated by $\left[\frac{\varpi(u)}{u}+\frac{1}{2}\varpi'(u)\right]\bar{r}2\pi$. Since the latter is integrable on $[\alpha,1]$ ($\int_\alpha^1\left[\frac{\varpi(u)}{u}+\frac{1}{2}\varpi'(u)\right]du\le\int_\alpha^1\left[\frac{1}{u}+\frac{1}{2}\varpi'(u)\right]du=\ln\frac{1}{\alpha}-\frac{1}{2}$) Dominated convergence means we can bring the limit inside the integral. Since $\lim_{n\to0}\tilde{t}\left(\psi n^{-1/2}u\right)\left(\psi n^{-1/2}u\right)^2=0$, the limit of the second terms is zero. ■

Claim A.3 *If* $\lim_{\delta\to\infty}\tilde{t}(\delta)\delta^4=0$, *then* $\kappa''(0)=0$

Proof. The second derivative of κ at zero is defined as $\kappa''(0) = \lim_{n\to 0} \frac{\kappa'(n)-\kappa'(0)}{n}$. $\lim_{\delta\to\infty} \tilde{t}(\delta)\delta^4 = 0$ implies $\lim_{\delta\to\infty} \tilde{t}(\delta)\delta^2 = 0$, so Claim A.2 gives $\kappa'(0) = \int_0^\infty \tilde{t}(\delta)2\pi\delta d\delta$. Using this along (18) gives

$$\begin{aligned}
\kappa''(0) &= \lim_{n\to 0} \frac{\frac{\psi^2}{n}\int_0^1 \left[\varpi(u) + \frac{1}{2}\varpi'(u)u\right]\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du - \int_0^\infty \tilde{t}\left(\delta\right)2\pi\delta d\delta}{n} \\
&= \lim_{n\to 0} \frac{\frac{\psi^2}{n}\int_\alpha^1 \left[\varpi(u) + \frac{1}{2}\varpi'(u)u\right]\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du}{n} + \frac{\frac{\psi^2}{n}\int_0^\alpha \tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du - \int_0^\infty \tilde{t}\left(\delta\right)2\pi\delta d\delta}{n}
\end{aligned}$$

We next show that each of the two terms is equal to zero. We first rearrange the first term and take the limit inside the integral using dominated convergence (the function $\left[-\left(\varpi(u) + \frac{1}{2}\varpi'(u)u\right)\right]\frac{2\pi u}{u^4}$ is integrable on the domain $u \in [\alpha, 1]$, in particular $\int_\alpha^1 \left[-\left(\varpi(u) + \frac{1}{2}\varpi'(u)u\right)\right]2\pi u du = 4\alpha$, and the fact that $\lim_{y\to\infty} \tilde{t}(y)y^4 = 0$ implies that $\tilde{t}(y)y^4$ has a uniform upper bound)

$$\begin{aligned}
&\lim_{n\to 0} \frac{\frac{\psi^2}{n}\int_\alpha^1 -\left[\varpi(u) + \frac{1}{2}\varpi'(u)u\right]\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du}{n} \\
&= \lim_{n\to 0} \frac{1}{\psi^2}\int_\alpha^1 -\left[\varpi(u) + \frac{1}{2}\varpi'(u)u\right]\tilde{t}\left(\psi n^{-1/2}u\right)\left(\psi n^{-1/2}u\right)^4 \frac{1}{u^4}2\pi u du \\
&= \frac{1}{\psi^2}\int_\alpha^1 -\left[\varpi(u) + \frac{1}{2}\varpi'(u)u\right]\left[\lim_{n\to 0}\tilde{t}\left(\psi n^{-1/2}u\right)\left(\psi n^{-1/2}u\right)^4\right]\frac{2\pi u}{u^4}du \\
&= \frac{1}{\psi^2}\int_\alpha^1 -\left[\varpi(u) + \frac{1}{2}\varpi'(u)u\right]\left[\lim_{y\to\infty}\tilde{t}(y)y^4\right]\frac{2\pi u}{u^4}du \\
&= \frac{4\alpha}{\psi^2}\lim_{y\to\infty}\tilde{t}\left(y\right)y^4 \\
&= 0
\end{aligned}$$

For the second term, we can change variables and use L'Hopital's rule.

$$\begin{aligned}
\lim_{n\to 0} \frac{\frac{\psi^2}{n}\int_0^\alpha \tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du - \int_0^\infty \tilde{t}\left(\delta\right)2\pi\delta d\delta}{n} &= \lim_{n\to 0} \frac{\int_0^{\alpha\psi n^{-1/2}} \tilde{t}\left(\delta\right)2\pi\delta d\delta - \int_0^\infty \tilde{t}\left(\delta\right)2\pi\delta d\delta}{n} \\
&= \lim_{n\to 0} \tilde{t}\left(\alpha\psi n^{-1/2}\right)2\pi\alpha\psi n^{-1/2}\left(-\frac{1}{2}\alpha\psi n^{-3/2}\right) \\
&= -\frac{\pi}{\left(\alpha\psi\right)^2}\lim_{n\to 0}\tilde{t}\left(\alpha\psi n^{-1/2}\right)\left(\alpha\psi n^{-1/2}\right)^4 \\
&= -\frac{\pi}{\left(\alpha\psi\right)^2}\lim_{y\to\infty}\tilde{t}(y)y^4 \\
&= 0
\end{aligned}$$

Together, these imply that $\kappa'(0) = 0$. ■

A.2 Proof of Main Proposition

Define $G(x)$ to be the integral of $T\left(\|s\|\right)^{1-\varepsilon}$ over a regular hexagon with area x centered at the origin. This will be an important function in characterizing optimal policy. Notice that $G(x) = \Delta^2 g\left(\frac{x}{\Delta^2}\right)$.[27] We begin by restating a well-known result from discrete geometry.

Theorem A.4 *(Theorem of L. Fejes Toth on sums of moments): Let $f : [0,\infty) \to \mathbb{R}$ be a nondecreasing function and let H be a convex 3,4,5, or 6-gon in $\mathbb{E}^2$. Then for any set of n points P in $\mathbb{E}^2$,*

$$\int_H \min\left\{f\left(\|x-p\|\right) : p \in P\right\} dx \geq n \int_{H_n} f\left(\|x\|\right) dx$$

where H_n is a regular hexagon in $\mathbb{E}^2$ with area $|H|/n$ and center at the origin.

We next apply this theorem to our context.

Lemma A.5 *For any h and any finite set of points $O_i \subset \mathcal{S}_i^k$,*

$$\int_{s\in\mathcal{S}_i^k} \max_{o\in O_i} T\left(\delta_{so}\right)^{1-\varepsilon} ds \leq |O_i|\, G\left(\frac{k^2}{|O_i|}\right)$$

Proof. Since $T(\delta)$ is strictly increasing in δ, $T(\delta)^{1-\varepsilon}$ is strictly decreasing. The theorem of L. Fejes Toth on sums of moments therefore implies that

$$\int_{s\in\mathcal{S}_i^k} \max_{o\in O_i} T\left(\delta_{so}\right)^{1-\varepsilon} ds = -\int_{s\in\mathcal{S}_i^k} \min_{o\in O_i} \left(-T\left(\delta_{so}\right)^{1-\varepsilon}\right) ds \leq -|O_i| \left(-G\left(\frac{k^2}{|O_i|}\right)\right) = |O_i|\, G\left(\frac{k^2}{|O_i|}\right)$$

■

Lemma A.6 *For any $k > 0, N \in \mathbb{N}_0$*

$$\sup_{O_i\subseteq\mathcal{S}_i^k || |O_i|=N} \int_{s\in\mathcal{S}_i} \max_{o\in O_i} T\left(\delta_{so}\right)^{1-\varepsilon} ds \geq NG\left(\rho(n)\frac{k^2}{N}\right)$$

where $\rho(n) = \left(1 + \frac{3^{3/4}}{\sqrt{2N}}\right)^{-2}$.

Proof. As in the proof of the lemma above, define $\psi \equiv 2^{1/2}3^{-3/4}$. The set $\mathcal{S}_i^k$ is a square with side length k. It is sufficient to show that for any non-negative integer N, one can fit N regular hexagons with area

[27] The definition of t along with the change of variables $\tilde{\delta} = \frac{\delta}{\Delta}$ imply

$$\begin{aligned} G(x) &= \int_0^{\psi\sqrt{x}} \varpi\left(\frac{\delta}{\psi\sqrt{x}}\right) t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon} 2\pi\delta d\delta = \Delta^2 \int_0^{\frac{1}{\Delta}\psi\sqrt{x}} \varpi\left(\frac{\tilde{\delta}}{\frac{1}{\Delta}\psi\sqrt{x}}\right) t\left(\tilde{\delta}\right)^{1-\varepsilon} 2\pi\tilde{\delta} d\tilde{\delta} \\ &= \Delta^2 g\left(\frac{x}{\Delta^2}\right) \end{aligned}$$

$\left(1+\frac{1}{\psi\sqrt{N}}\right)^{-2}\frac{k^2}{N}$ inside the square $\mathcal{S}_i^k$ as this would constitute a particular O_i choice. Using our hexagon area expression from Lemma 2, each of these hexagons would have a side length $l=\frac{\psi}{\sqrt{N}}\frac{1}{1+\psi^{-1}N^{-1/2}}k$. Since regular hexagons can form a regular tiling of the plane, we can consider hexagons each with side length l and tiling with $c=\left\lceil\psi\sqrt{3}\sqrt{N}\right\rceil$ columns and $r=\left\lceil\frac{1}{\psi\sqrt{3}}\sqrt{N}\right\rceil$ rows, where $\lceil x\rceil$ denotes the smallest integer weakly larger than x. Our proposed lattice has total width of $\left(\frac{3}{2}c+\frac{1}{2}\right)l$ and total height weakly less than $(2r+1)\sqrt{l^2-(l/2)^2}=\left(\sqrt{3}r+\frac{\sqrt{3}}{2}\right)l$ (with equality if there is more than one column). We first show that that total width is smaller than k, i.e., $\left(\frac{3}{2}c+\frac{1}{2}\right)l\leq k$. To see this, we have

$$\begin{aligned}\left(\frac{3}{2}c+\frac{1}{2}\right)l &= \left(\frac{3}{2}\left\lceil\psi\sqrt{3}\sqrt{N}\right\rceil+\frac{1}{2}\right)\frac{\psi}{\sqrt{N}}\frac{1}{1+\psi^{-1}N^{-1/2}}k\\ &\leq \left(\frac{3}{2}\left(\psi\sqrt{3}\sqrt{N}+1\right)+\frac{1}{2}\right)\frac{\psi}{\sqrt{N}}\frac{1}{1+\psi^{-1}N^{-1/2}}k\\ &= \frac{1+\frac{2\psi}{\sqrt{N}}}{1+\psi^{-1}N^{-1/2}}k\\ &\leq k\end{aligned}$$

where the last step follows because $4\leq 3^{3/2}$ implies $2\psi=\frac{2}{2^{-1/2}3^{3/4}}\leq 3^{3/4}2^{-1/2}=\frac{1}{\psi}$.

We next show that the total height is less than k, i.e., $\left(\sqrt{3}r+\frac{\sqrt{3}}{2}\right)l\leq k$. To see this, we have

$$\begin{aligned}\left(\sqrt{3}r+\frac{\sqrt{3}}{2}\right)l &= \left(\sqrt{3}\left\lceil\frac{1}{\psi\sqrt{3}}\sqrt{N}\right\rceil+\frac{\sqrt{3}}{2}\right)\frac{\psi}{\sqrt{N}}\frac{1}{1+\psi^{-1}N^{-1/2}}k\\ &\leq \left(\sqrt{3}\left(\frac{1}{\psi\sqrt{3}}\sqrt{N}+1\right)+\frac{\sqrt{3}}{2}\right)\frac{\psi}{\sqrt{N}}\frac{1}{1+\psi^{-1}N^{-1/2}}k\\ &= \frac{1+\frac{1}{\psi\sqrt{N}}}{1+\psi^{-1}N^{-1/2}}k=k\end{aligned}$$

Finally, we note that such a lattice contains $cr=\left\lceil 2^{1/2}3^{-1/4}\sqrt{N}\right\rceil\left\lceil 2^{-1/2}3^{1/4}\sqrt{N}\right\rceil\geq N$ regular hexagons. It follows that N regular hexagons each with area $\left(1+\frac{1}{\psi\sqrt{N}}\right)^{-2}\frac{k^2}{N}$ fit inside the square $\mathcal{S}_i^k$. ■

Claim A.7 *For any k, Δ, $\bar{\pi}_j^{k\Delta}\geq\pi_j^{\Delta}$ where*

$$\bar{\pi}_j^{k\Delta}\equiv\sup_{\{N_i\geq 0\}}\sum_i -N_i\underline{R}_i^k\xi+Z\left(q_j,|O|\right)^{\varepsilon-1}\sum_i\bar{D}_i^k\bar{b}_i^kN_iG\left(\frac{k^2}{N_i}\right)+Z\left(q_j,0\right)^{\epsilon-1}\bar{D}\bar{b}\frac{1}{k}\int_0^k T\left(\delta\right)^{1-\varepsilon}d\delta$$

Proof. For any set of plants O, let O_i^k be the subset that are in square i. We begin with:

$$\begin{aligned}
\int_s D_s \max_{o\in O}\left\{b_o T\left(\delta_{so}\right)^{1-\varepsilon}\right\} ds &= \sum_{i\in I^k}\int_{s\in\mathcal{S}_i^k} D_s \max_{o\in O}\left\{b_o T\left(\delta_{so}\right)^{1-\varepsilon}\right\} ds \\
&\leq \sum_{i\in I^k}\int_{s\in\mathcal{S}_i^k} D_s \sum_{i'\in I^k}\max_{o\in O_{i'}^k}\left\{b_o T\left(\delta_{so}\right)^{1-\varepsilon}\right\} ds \\
&\leq \sum_{i\in I^k}\int_{s\in\mathcal{S}_i^k} \bar{D}_i^k \sum_{i'\in I^k}\bar{b}_{i'}^k \max_{o\in O_{i'}^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds \\
&= \sum_{i\in I^k}\bar{D}_i^k\bar{b}_i^k\int_{s\in\mathcal{S}_i^k}\max_{o\in O_i^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds + \sum_{i\in I^k}\sum_{i'\neq i}\bar{D}_i^k\bar{b}_{i'}^k\int_{s\in\mathcal{S}_i^k}\max_{o\in O_{i'}^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds
\end{aligned}$$

We can bound the first term using the previous lemma. To bound the second term, note that

$$\int_{s\in\mathcal{S}_i^k}\max_{o\in O_{i'}^k} T(\delta_{so})^{1-\varepsilon} ds \leq \int_{s\in\mathcal{S}_i^k}\max_{o\in\mathcal{S}_{i'}^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds.$$

The term $\int_{s\in\mathcal{S}_i^k}\max_{o\in O_{i'}^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds$ is maximized if $\mathcal{S}_i^k$, $\mathcal{S}_{i'}^k$ are contiguous, in which case,

$$\int_{s\in\mathcal{S}_i^k}\max_{o\in\mathcal{S}_{i'}^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds = k\int_0^k T(\delta)^{1-\varepsilon} d\delta$$

In addition, $\sum_i\sum_{i'\neq i}\bar{D}_i^k\bar{b}_{i'}^k \leq \frac{1}{k^2}\bar{D}\bar{b}$, so that $\sum_i\sum_{i'\neq i}\bar{D}_i^k\bar{b}_{i'}^k\int_{s\in\mathcal{S}_i^k}\max_{o\in O_{i'}^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds \leq \bar{D}\bar{b}\frac{1}{k}\int_0^k T(\delta)^{1-\varepsilon} d\delta$
Together, these imply

$$\int_s D_s \max_{o\in O}\left\{b_o T\left(\delta_{so}\right)^{1-\varepsilon}\right\} ds \leq \sum_i \bar{D}_i^k\bar{b}_i^k\int_{s\in\mathcal{S}_i^k}\left|O_i^k\right| G\left(\frac{k^2}{\left|O_i^k\right|}\right) + \bar{D}\bar{b}\frac{1}{k}\int_0^k T(\delta)^{1-\varepsilon} d\delta$$

Similarly,

$$\sum_{o\in O} -R_o\xi = \sum_i\sum_{o\in O_i^k} -R_o\xi \leq \sum_i\sum_{o\in O_i^k} -\underline{R}_i^k\xi = \sum_i -\left|O_i^k\right|\underline{R}_i^k\xi$$

Together, these imply that

$$\begin{aligned}
\pi_j^{\Delta} &= \sup_{O} \sum_{o \in O} -R_o \xi + Z\left(q_j, |O|\right)^{\varepsilon-1} \int_s D_s \max_{o \in O} \left\{ b_o T\left(\delta_{so}\right)^{1-\varepsilon} \right\} ds \\
&\leq \sup_{O} \sum_i - \left| O_i^k \right| \underline{R}_i^k \xi + Z\left(q_j, |O|\right)^{\varepsilon-1} \left(\sum_i \bar{D}_i^k \bar{b}_i^k \left| O_i^k \right| G\left(\frac{k^2}{\left| O_i^k \right|} \right) + \bar{D}\bar{b} \frac{1}{k} \int_0^k T(\delta)^{1-\varepsilon} d\delta \right) \\
&\leq \sup_{\{N_i \in \mathbb{N}_0\}} \sum_i -N_i \underline{R}_i^k \xi + Z\left(q_j, |O|\right)^{\varepsilon-1} \sum_i \bar{D}_i^k \bar{b}_i^k N_i G\left(\frac{k^2}{N_i} \right) + Z\left(q_j, 0\right)^{\epsilon-1} \bar{D}\bar{b} \frac{1}{k} \int_0^k T(\delta)^{1-\varepsilon} d\delta \\
&\leq \sup_{\{N_i \geq 0\}} \sum_i -N_i \underline{R}_i^k \xi + Z\left(q_j, |O|\right)^{\varepsilon-1} \sum_i \bar{D}_i^k \bar{b}_i^k N_i G\left(\frac{k^2}{N_i} \right) + Z\left(q_j, 0\right)^{\epsilon-1} \bar{D}\bar{b} \frac{1}{k} \int_0^k T\left(\delta\right)^{1-\varepsilon} d\delta \\
&= \bar{\pi}_j^{k\Delta}
\end{aligned}$$

where we use that Z is decreasing in N and slightly abuse notation so that N_i for a particular firm is choice of number of plants in $\mathcal{S}_i^k$. ■

Claim A.8 *In the limit as* $\Delta \to 0$*,*

$$\lim_{\Delta \to 0} \bar{\pi}^{k\Delta} \leq \bar{\pi}^k \equiv \sup_{n \geq 0} \int \left\{ -n_s \underline{R}_s^k + z\left(q_j, \int n_{\tilde{s}} d\tilde{s} \right)^{\varepsilon-1} \bar{D}_s^k \bar{b}_s^k \kappa\left(n_s\right) \right\} ds$$

Proof. Replace and use $n_i = \frac{\Delta^2 N_i}{k^2}$

$$\begin{aligned}
\bar{\pi}_j^{k\Delta} &= \sup_{\{N_i \geq 0\}} \sum_i \left\{ -N_i \underline{R}_i^k \xi + Z\left(q_j, N\right)^{\varepsilon-1} \bar{D}_i^k \bar{b}_i^k N_i G\left(\frac{k^2}{N_i} \right) \right\} + Z\left(q_j, 0\right)^{\varepsilon-1} \bar{D}\bar{b} \frac{1}{k} \int_0^k T(\delta)^{1-\varepsilon} d\delta \\
&= \sup_{\{N_i \geq 0\}} \sum_i \left\{ -N_i \underline{R}_i^k \Delta^2 + z\left(q_j, \Delta^2 N\right)^{\varepsilon-1} \bar{D}_i^k \bar{b}_i^k N_i \Delta^2 g\left(\frac{1}{\Delta^2} \frac{k^2}{N_i} \right) \right\} + Z\left(q_j, 0\right)^{\varepsilon-1} \bar{D}\bar{b} \frac{1}{k} \int_0^k t \left(\frac{\delta}{\Delta} \right)^{1-\varepsilon} d\delta \\
&= \sup_{\{n_i \geq 0\}} k^2 \sum_i \left\{ -n_i \underline{R}_i^k + z\left(q_j, k^2 \sum_{\tilde{i}} n_{\tilde{i}} \right)^{\varepsilon-1} \bar{D}_i^k \bar{b}_i^k n_i g\left(\frac{1}{n_i} \right) \right\} + Z\left(q_j, 0\right)^{\varepsilon-1} \bar{D}\bar{b} \frac{1}{k} \int_0^k t \left(\frac{\delta}{\Delta} \right)^{1-\varepsilon} d\delta
\end{aligned}$$

Taking the limit gives

$$\begin{aligned}
\lim_{\Delta\to 0}\bar{\pi}_j^{k\Delta} &= \lim_{\Delta\to 0}\sup_{\{n_i\geq 0\}} k^2\sum_i\left\{-n_i\underline{R}_i^k + z\left(q_j, k^2\sum_{\tilde{\imath}} n_{\tilde{\imath}}\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k n_i g\left(\frac{1}{n_i}\right)\right\} + \\
&+ \lim_{\Delta\to 0} z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}\frac{1}{k}\int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}d\delta \\
&= \sup_{\{n_i\geq 0\}} k^2\sum_i\left\{-n_i\underline{R}_i^k + z\left(q_j, k^2\sum_{\tilde{\imath}} n_{\tilde{\imath}}\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k n_i g\left(\frac{1}{n_i}\right)\right\} \\
&= \sup_{\{n_i\geq 0\}} k^2\sum_i\left\{-n_i\underline{R}_i^k + z\left(q_j, k^2\sum_{\tilde{\imath}} n_{\tilde{\imath}}\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k\kappa\left(n_i\right)\right\}
\end{aligned}$$

where we used the assumption that $t(\delta)$ diverges as $\delta\to\infty$. Let $\mathcal{N}^k$ be the set of strategies in which n_s is constant for all $s\in\mathcal{S}_i^k$. Then we can write

$$\begin{aligned}
\lim_{\Delta\to 0}\bar{\pi}_j^{k\Delta} &= \sup_{n\in\mathcal{N}^k}\int\left\{-n_s\underline{R}_s^k + z\left(q_j, \int n_{\tilde{s}}d\tilde{s}\right)^{\varepsilon-1}\bar{D}_s^k\bar{b}_s^k\kappa\left(n_s\right)\right\}ds \\
&\leq \sup_{n\geq 0}\int\left\{-n_s\underline{R}_s^k + z\left(q_j, \int n_{\tilde{s}}d\tilde{s}\right)^{\varepsilon-1}\bar{D}_s^k\bar{b}_s^k\kappa\left(n_s\right)\right\}ds \\
&= \bar{\pi}^k
\end{aligned}$$

■

We next show that for each j, it is without loss of generality to impose a uniform upper bound on the density of plants.

Lemma A.9 *Define $\bar{n}_j$ to satisfy $\underline{R} = z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}\kappa'(\bar{n}_j)$.*

$$\bar{\pi}_j^k = \sup_{n\in[0,\bar{n}_j]}\int\left\{-n_s\underline{R}_s^k + z\left(q_j, \int_{\tilde{s}} n_{\tilde{s}}d\tilde{s}\right)^{\varepsilon-1}\bar{D}_s^k\bar{b}_s^k\kappa\left(n_s\right)\right\}ds$$

Proof. We show that it is without loss to restrict the strategies to the set $n\in[0,\bar{n}_j]$. Restricting the set of strategies yields a weakly lower payoff. To show the opposite inequality, for any strategy, let N^+ be the subset of $\mathcal{S}$ for which $n_s>\bar{n}_j$. We will show that the alternative strategy in which

$$\tilde{n} = \begin{cases} n_s & s\notin N^+ \\ \bar{n}_j & s\in N^+ \end{cases}$$

would give a weakly higher payoff, Consider any profile R, D, b such that $R_s\geq\underline{R}$ and $D_s\leq\bar{D}$ and $b_s\leq\bar{b}$.

For shorthand, we express $z_j = z\left(q_j, \int_s n_s ds\right)$ and $\tilde{z}_j = z\left(q_j, \int_s \tilde{n}_s ds\right)$.

$$\begin{aligned}
\Pi_j(n) - \Pi_j(\tilde{n}) &= \int \left\{-n_s R_s + z_j^{\varepsilon-1} D_s b_s h \kappa(n_s)\right\} ds - \Pi(\tilde{n}) \\
&\leq \int \left\{-n_s R_s + \tilde{z}_j^{\varepsilon-1} D_s b_s \kappa(n_s)\right\} ds - \Pi(\tilde{n}) \\
&= \int_{s\in N^+} \left\{-n_s R_s + \tilde{z}_j^{\varepsilon-1} D_s b_s \kappa(n_s)\right\} ds - \int_{s\in N^+} \left\{-\tilde{n}_s R_s + \tilde{z}_j^{\varepsilon-1} D_s b_s \kappa(\tilde{n}_s)\right\} ds \\
&= \int_{s\in N^+} \left\{-n_s R_s + \tilde{z}_j^{\varepsilon-1} D_s b_s \kappa(n_s)\right\} ds - \int_{s\in N^+} \left\{-\bar{n}_s R_s + \tilde{z}_j^{\varepsilon-1} D_s b_s \kappa_s(\bar{n})\right\} ds \\
&= \int_{s\in N^+} \left\{-(n_s - \bar{n}_j) R_s + \tilde{z}_j^{\varepsilon-1} D_s b_s \left[\kappa(n_s) - \kappa(\bar{n})\right]\right\} ds
\end{aligned}$$

The concavity of κ implies $\kappa(n_s) \leq \kappa(\bar{n}_j) + \kappa'(\bar{n}_j)(n_s - \bar{n}_j)$, so that $\kappa(n_s) - \kappa(\bar{n}_j) \leq \kappa'(\bar{n}_j)(n_s - \bar{n}_j) = \frac{\underline{R} a}{z(q_j,0)^{\varepsilon-1} \bar{D}\bar{b}}(n_s - \bar{n}_j)$. Plugging this in gives

$$\begin{aligned}
\Pi_j(n) - \Pi_j(\tilde{n}) &\leq \int_{s\in N^+} \left\{-(n_s - \bar{n}_j) R_s + \tilde{z}_j^{\varepsilon-1} D_s b_s \frac{\underline{R}}{z(q_j,0)^{\varepsilon-1} \bar{D}\bar{b}}(n_s - \bar{n}_j)\right\} ds \\
&= \int_{s\in N^+} \left\{-1 + \frac{\tilde{z}_j^{\varepsilon-1} D_s b_s}{z(q_j,0)^{\varepsilon-1} \bar{D}\bar{b}} \frac{\underline{R}}{R_s}\right\} R_s (n_s - \bar{n}_j) ds \\
&\leq 0
\end{aligned}$$

■

Claim A.10

$$\pi_j \leq \sup_{n\geq 0} \int \left\{-n_s R_s + z\left(q_j, \int_{\tilde{s}} n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} D_s b_s \kappa(n_s)\right\} ds$$

Proof. For any strategy n, define $\bar{\Pi}^k(n)$ and $\Pi(n)$ as

$$\begin{aligned}
\bar{\Pi}^k(n) &= \int \left\{-n_s \underline{R}_s^k + z\left(q, \int_{\tilde{s}} n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} \bar{D}_s^k \bar{b}_s^k \kappa(n_s)\right\} ds \\
\Pi(n) &= \int \left\{-n_s R_s + z\left(q_j, \int_{\tilde{s}} n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} D_s b_s \kappa(n_s)\right\} ds
\end{aligned}$$

Since R, D, and b are continuous on a compact space, they are uniformly continuous. This implies that for any $\varphi > 0$, there is an η small enough so that $h < \eta$ implies both $\left|\underline{R}_s^k - R_s\right| \leq \varphi$, and $\left|\bar{D}_s^k \bar{b}_s^k - D_s b_s\right| \leq \varphi$.

With that, for any $n \in [0, \bar{n}_j]$,

$$\begin{aligned}
\left|\bar{\Pi}^k(n) - \Pi(n)\right| &= \left|\int \left\{-n_s\left(\underline{R}^k_s - R_s\right) + z\left(q_j, \int_{\tilde{s}} n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1}\left[\bar{D}^k_s \bar{b}^k_s - D_s b_s\right]\kappa(n_s)\right\} ds\right| \\
&\leq \int \left\{n_s\left|\underline{R}^k_s - R_s\right| + z\left(q_j, \int_{\tilde{s}} n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1}\left|\bar{D}^k_s \bar{b}^k_s - D_s b_s\right|\kappa(n_s)\right\} ds \\
&\leq \int \left\{n_s\varphi + z\left(q_j, \int_{\tilde{s}} n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1}\varphi\kappa(n_s)\right\} ds \\
&\leq \int \left\{\bar{n}_j\varphi + z\left(q_j, 0\right)^{\varepsilon-1}\varphi\right\} ds \\
&\leq \varphi \int \left\{\bar{n}_j + z\left(q_j, 0\right)^{\varepsilon-1}\right\} ds
\end{aligned}$$

Therefore $\bar{\Pi}^k(\cdot)$ is uniformly convergent on the domain $n \in [0, \bar{n}_j]$ as $k \to 0$.. Therefore

$$\lim_{k\to 0} \sup_{n\in[0,\bar{n}_j]} \bar{\Pi}^k(n) = \sup_{n\in[0,\bar{n}_j]} \lim_{k\to 0} \bar{\Pi}^k(n)$$

In other words, we have $\pi_j \leq \bar{\pi}^k_j$ for all k, so taking the limit of both sides yields

$$\pi_j \leq \lim_{k\to 0} \bar{\pi}^k_j = \lim_{k\to 0} \sup_{n\in[0,\bar{n}_j]} \bar{\Pi}^k(n) = \sup_{n\in[0,\bar{n}_j]} \lim_{k\to 0} \bar{\Pi}^k(n) = \sup_{n\in[0,\bar{n}_j]} \Pi(n) \leq \sup_{n\geq 0} \Pi(n)$$

■

We next bound the payoff from below.

Claim A.11

$$\pi^{\Delta}_j \geq \underline{\pi}^{k\Delta}_j \equiv \sup_{\{N_i\in\mathbb{N}_0\}} \sum_i -N_i \bar{R}^k_i \xi + Z\left(q_j, \sum_{\tilde{i}} N_{\tilde{i}}\right)^{\varepsilon-1} \sum_i \underline{D}^k_i \underline{b}^k_i N_i G\left(\rho(N_i)\frac{k^2}{N_i}\right)$$

Proof. Begin with

$$\begin{aligned}
\int_s D_s \max_{o\in O}\left\{b_o T(\delta_{so})^{1-\varepsilon}\right\} ds &= \sum_i \int_{s\in\mathcal{S}^k_i} D_s \max_{o\in O}\left\{b_o T(\delta_{so})^{1-\varepsilon}\right\} ds \\
&\geq \sum_i \int_{s\in\mathcal{S}^k_i} D_s \max_{o\in O^k_i}\left\{b_o T(\delta_{so})^{1-\varepsilon}\right\} ds \\
&\geq \sum_i \underline{D}^k_i \underline{b}^k_i \int_{s\in\mathcal{S}^k_i} \max_{o\in O^k_i} T(\delta_{so})^{1-\varepsilon} ds
\end{aligned}$$

Similarly,

$$\sum_{o\in O} -R_o\xi = \sum_i \sum_{o\in O_i^k} -R_o\xi \geq \sum_i \sum_{o\in O_i^k} -\bar{R}_i^k\xi = \sum_i -\left|O_i^k\right| \bar{R}_i^k\xi$$

Together, these yield a lower bound for π_j^{Δ}

$$\begin{aligned}
\pi_j^{\Delta} &= \sup_O \sum_{o\in O} -R_o\xi + Z\left(q_j,|O|\right)^{\varepsilon-1}\int_s D_s \max_{o\in O}\left\{b_o T\left(\delta_{so}\right)^{1-\varepsilon}\right\} ds \\
&\geq \sup_O \sum_i -\left|O_i^k\right| \bar{R}_i^k\xi + Z\left(q_j,|O|\right)^{\varepsilon-1}\sum_i \underline{D}_i^k \underline{b}_i^k \int_{s\in\mathcal{S}_i^k} \max_{o\in O_i} T\left(\delta_{so}\right)^{1-\varepsilon} ds \\
&= \sup_{\{N_i\in\mathbb{N}_0\}} \sup_{\left\{O_i^k\subset\mathcal{S}_i^k \middle| \left|O_i^k\right|=N_i\right\}} \sum_i -\left|O_i^k\right| \bar{R}_i^k\xi + Z\left(q_j,|O|\right)^{\varepsilon-1}\sum_i \underline{D}_i^k \underline{b}_i^k \int_{s\in\mathcal{S}_i^k} \max_{o\in O_i^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds \\
&= \sup_{\{N_i\in\mathbb{N}_0\}} \sup_{\left\{O_i\subset\mathcal{S}_i^k \middle| \left|O_i^k\right|=N_i\right\}} \sum_i -N_i\bar{R}_i^k\xi + Z\left(q,\sum_i N_i\right)^{\varepsilon-1}\sum_i \underline{D}_i^k \underline{b}_i^k \int_{s\in\mathcal{S}_i^k} \max_{o\in O_i^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds \\
&= \sup_{\{N_i\in\mathbb{N}_0\}} \sum_i -N_i\bar{R}_i^k\xi + Z\left(q_j,\sum_i N_i\right)^{\varepsilon-1}\sum_i \underline{D}_i^k \underline{b}_i^k \sup_{\left\{O_i^k\subset\mathcal{S}_i^k \middle| \left|O_i^k\right|=N_i\right\}} \int_{s\in\mathcal{S}_i^k} \max_{o\in O_i^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds \\
&\geq \sup_{\{N_i\in\mathbb{N}_0\}} \sum_i -N_i\bar{R}_i^k\xi + Z\left(q_j,\sum_i N_i\right)^{\varepsilon-1}\sum_i \underline{D}_i^k \underline{b}_i^k N_i G\left(\rho\left(N_i\right)\frac{k^2}{N_i}\right) \\
&= \underline{\pi}_j^{k\Delta}
\end{aligned}$$

■

Claim A.12 *For any k, in the limit as $\Delta\to 0$,*

$$\lim_{\Delta\to 0} \underline{\pi}_j^{k\Delta} \geq \underline{\pi}^k \equiv \sup_{\{n_s\geq 0\}} \int_s \left\{-n_s\bar{R}_s^k + z\left(q_j,\int n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} \underline{D}_s^k \underline{b}_s^k \kappa\left(n_s\right)\right\} ds$$

Proof. Replace and use $n_i = \frac{\Delta^2 N_i}{k^2}$

$$\begin{aligned}
\underline{\pi}_j^{k\Delta} &= \sup_{\{N_i \in \mathbb{N}_0\}} \sum_i -N_i \bar{R}_i^k \xi + Z\left(q_j, \sum_i N_i\right)^{\varepsilon-1} \sum_i \underline{D}_i^k \underline{b}_i^k N_i G\left(\rho\left(N_i\right) \frac{k^2}{N_i}\right) \\
&= \sup_{\{N_i \geq 0\}} \sum_i -\lceil N_i \rceil \bar{R}_i^k \xi + Z\left(q_j, \sum_i \lceil N_i \rceil\right)^{\varepsilon-1} \sum_i \underline{D}_i^k \underline{b}_i^k \lceil N_i \rceil G\left(\rho\left(\lceil N_i \rceil\right) \frac{k^2}{\lceil N_i \rceil}\right) \\
&= \sup_{\{N_i \geq 0\}} \sum_i -\lceil N_i \rceil \bar{R}_i^k \Delta^2 + z\left(q_j, \Delta^2 \sum_i \lceil N_i \rceil\right)^{\varepsilon-1} \sum_i \underline{D}_i^k \underline{b}_i^k \lceil N_i \rceil \Delta^2 g\left(\frac{1}{\Delta^2}\rho\left(\lceil N_i \rceil\right) \frac{k^2}{\lceil N_i \rceil}\right) \\
&= \sup_{\{n_i \geq 0\}} \sum_i -\left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil \bar{R}_i^k \Delta^2 + z\left(q_j, \Delta^2 \sum_i \left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil\right)^{\varepsilon-1} \sum_i \underline{D}_i^k \underline{b}_i^k \left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil \Delta^2 g\left(\frac{1}{\Delta^2}\rho\left(\left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil\right) \frac{k^2}{\left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil}\right)
\end{aligned}$$

Next, we use the fact that $\liminf_{\Delta\to 0} \sup_{\{n\geq 0\}} f\left(n, \Delta\right) \geq \sup_{\{n\geq 0\}} \liminf_{\Delta\to 0} f\left(n, \Delta\right)$[28] along with $\lim_{\Delta\to 0} \Delta^2 \left\lceil \frac{h^2}{\Delta^2} n_i \right\rceil = h^2 n_i$ and $\lim_{N\to\infty} \rho(n) = 1$ to get

$$\begin{aligned}
\lim_{\Delta\to 0} \underline{\pi}_j^{k\Delta} &\geq \sup_{\{n_i \geq 0\}} \lim_{\Delta\to 0} \sum_i -\left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil \bar{R}_i^k \Delta^2 + \\
&+ z\left(q_j, \Delta^2 \sum_i \left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil\right)^{\varepsilon-1} \sum_i \underline{D}_i^k \underline{b}_i^k \left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil \Delta^2 g\left(\frac{1}{\Delta^2}\rho\left(\left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil\right) \frac{k^2}{\left\lceil \frac{k^2}{\Delta^2} n_i \right\rceil}\right) \\
&= \sup_{\{n_i \geq 0\}} \sum_i -k^2 n_i \bar{R}_i^k + z\left(q_j, \sum_i k^2 n_i\right)^{\varepsilon-1} \sum_i \underline{D}_i^k \underline{b}_i^k k^2 n_i g\left(\frac{1}{n_i}\right) \\
&= \sup_{\{n_i \geq 0\}} k^2 \left\{ \sum_i -n_i \bar{R}_i^k + z\left(q_j, k^2 \sum_i n_i\right)^{\varepsilon-1} \underline{D}_i^k \underline{b}_i^k \kappa\left(n_i\right) \right\}
\end{aligned}$$

Since $\kappa(n)$ is strictly concave, Jensen's inequality implies that

$$\sup_{n_s} \int_{s\in\mathcal{S}_i^k} \kappa\left(n_s\right) ds \text{ subject to } \int_{s\in\mathcal{S}_i^k} n_s ds \leq n_i$$

[28] Quick proof: For any n_0, Δ we have $f\left(n_0, \Delta\right) \leq \sup_n f\left(n, \Delta\right)$. Taking limits preserves inequalities, so that $\liminf_{\Delta\to 0} f\left(n_0, \Delta\right) \leq \liminf_{\Delta\to 0} \sup_n f\left(n, \Delta\right)$. The conclusion follows from taking sup of both sides with respect to n_0.

is maximized for $n_s = \frac{n_i}{|\mathcal{S}_i^k|}$, i.e., n_s is constant. This means

$$\begin{aligned}
\lim_{\Delta\to 0} \underline{\pi}_j^{k\Delta} &\geq \sup_{\{n_i\geq 0\}} k^2 \left\{ \sum_i -n_i \bar{R}_i^k + z\left(q_j, k^2 \sum_i n_i\right)^{\varepsilon-1} \underline{D}_i^k \underline{b}_i^k \kappa\left(n_i\right) \right\} \\
&= \sup_{\{n_s\geq 0\}} \int_s \left\{ -n_s \bar{R}_s^k + z\left(q_j, \int n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} \underline{D}_s^k \underline{b}_s^k \kappa\left(n_s\right) \right\} ds \\
&= \underline{\pi}_j^k
\end{aligned}$$

■

Claim A.13

$$\pi_j \geq \sup_{\{n_i\geq 0\}} \int \left\{ -n_s R_s + z\left(q_j, \int n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} D_s b_s \kappa\left(n_s\right) \right\} ds$$

Proof. We can again use the fact that $\liminf_{h\to 0} \sup_{\{n\geq 0\}} f(n,h) \geq \sup_{\{n\geq 0\}} \liminf_{h\to 0} f(n,h)$ to write as $k \to 0$

$$\begin{aligned}
\pi_j &\geq \liminf_{k\to 0} \pi_j^{k\Delta} \geq \liminf_{k\to 0} \underline{\pi}_j^k \\
&\geq \sup_{\{n_s\geq 0\}} \liminf_{k\to 0} \int_s \left\{ -n_s \bar{R}_s^k + z\left(q_j, \int n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} \underline{D}_s^k \underline{b}_s^k \kappa\left(n_s\right) \right\} ds \\
&= \sup_{\{n_s\geq 0\}} \int_s \left\{ -n_s R_s + z\left(q_j, \int n_{\tilde{s}} d\tilde{s}\right)^{\varepsilon-1} D_s b_s \kappa\left(n_s\right) \right\} ds
\end{aligned}$$

■

A.3 Aggregation

A.3.1 The Local Price Index

The price that firm j sets in location s is

$$p_{js} = \frac{\varepsilon}{\varepsilon - 1} \min_{o\in O_j} \left\{ \frac{W_o T\left(\delta_{so}\right)}{B_o Z\left(q, N_j\right)} \right\}$$

In a small enough neighborhood of location s, economic activity is locally uniform. Thus each firm will choose to have catchment areas that are locally uniform regular hexagons. Among firms with the same effective productivity Z, the pattern of plant locations will be the same up to translation. These translations are such that if we integrate across such firms, the total measure of plants at each point will be uniform. An implications is that, for consumers in location s and firms with effective productivity Z, the fraction of those firms that have plants closer than distance δ to those consumers is the same as the fraction of locations in

such a firm's catchment area that are closer than distance δ to the plant at the center of the catchment area.

Given this we now derive an expression for the ideal price index at a location. As in the proof of the main proposition, we will proceed by dividing the economy into $k \times k$ squares in which economic activity is uniform, taking the limit as $\Delta \to 0$, and then taking the limit as $k \to 0$. We ignore boundary issues because these will disappear when we take the limit as $\Delta \to 0$.

The ideal price index at location s satisfies $P_s^{1-\varepsilon} = \int p_{js}^{1-\varepsilon} dj$. Consider a $k \times k$ square with uniform local economic activity, so that catchment areas are uniform hexagons. We can compute the local ideal price index at any point in that $k \times k$ square by integrating over all firms in the economy.

Let N_{ji} be the number of plants that firm j places in the square. Then for each plant, the distance to the furthest point in the catchment area is $\psi\sqrt{k^2/N_{ji}}$, and among points that are distance δ from the plant, the fraction $\varpi\left(\frac{\delta}{\psi\sqrt{k^2/N_{ji}}}\right)$ are in the plant's catchment area (the remainder are served by other plants).[29] The ideal price index can therefore be expressed as

$$\left(P_s^{k\Delta}\right)^{1-\varepsilon} = \int \frac{\int_0^{\psi\sqrt{k^2/N_{ji}}} \varpi\left(\frac{\delta}{\psi\sqrt{k^2/N_{ji}}}\right) \left[\frac{\varepsilon}{\varepsilon-1}\frac{W_s T(\delta)}{B_s Z(q,N_j)}\right]^{1-\varepsilon} 2\pi\delta d\delta}{\int_0^{\psi\sqrt{k^2/N_{ji}}} \varpi\left(\frac{\delta}{\psi\sqrt{k^2/N_{ji}}}\right) 2\pi\delta d\delta} dj$$

Using $T(\delta) = t\left(\frac{\delta}{\Delta}\right)$ and $Z\left(q,N\right) \equiv z\left(q,\Delta^2 N\right)^{\varepsilon-1}$ gives

$$\left(P_s^{k\Delta}\right)^{1-\varepsilon} = \int \frac{\int_0^{\psi\sqrt{k^2/N_{ji}}} \varpi\left(\frac{\delta}{\psi\sqrt{k^2/N_{ji}}}\right) \left[\frac{\frac{\varepsilon}{\varepsilon-1} W_s t\left(\frac{\delta}{\Delta}\right)}{B_s z(q,\Delta^2 N_j)}\right]^{1-\varepsilon} 2\pi\delta d\delta}{\int_0^{\psi\sqrt{k^2/N_{ji}}} \varpi\left(\frac{\delta}{\psi\sqrt{k^2/N_{ji}}}\right) 2\pi\delta d\delta} dj$$

Using $n_{ji} = \frac{\Delta^2 N_{ji}}{k^2}$ and using a change of variables gives

$$\left(P_s^{k\Delta}\right)^{1-\varepsilon} = \left(\frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s}\right)^{1-\varepsilon} \int z_j^{\varepsilon-1} \frac{\int_0^{\psi n_{ji}^{-1/2}} \varpi\left(\frac{\tilde{\delta}}{\psi n_{ji}^{-1/2}}\right) t\left(\tilde{\delta}\right)^{1-\varepsilon} 2\pi\tilde{\delta}d\tilde{\delta}}{\int_0^{\psi n_{ji}^{-1/2}} \varpi\left(\frac{\tilde{\delta}}{\psi n_{ji}^{-1/2}}\right) 2\pi\tilde{\delta}d\tilde{\delta}} dj$$

[29] Recall that $\varpi(x)$ is defined as that fraction of a circle of radius x that intersects with the interior of a hexagon with side length 1.

Taking the limits as $\Delta \to 0$ and $k \to 0$ gives

$$\begin{aligned} P_s^{1-\varepsilon} &= \left(\frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s}\right)^{1-\varepsilon} \int z_j^{\varepsilon-1} \frac{\int_0^{\psi n_{js}^{-1/2}} \varpi\left(\frac{\tilde{\delta}}{\psi n_{js}^{-1/2}}\right) t\left(\tilde{\delta}\right)^{1-\varepsilon} 2\pi\tilde{\delta}d\tilde{\delta}}{\int_0^{\psi n_{js}^{-1/2}} \varpi\left(\frac{\tilde{\delta}}{\psi n_{js}^{-1/2}}\right) 2\pi\tilde{\delta}d\tilde{\delta}} dj \\ &= \left(\frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s}\right)^{1-\varepsilon} \int z_j^{\varepsilon-1} n \int_0^{\psi n_{js}^{-1/2}} \varpi\left(\frac{\tilde{\delta}}{\psi n_{js}^{-1/2}}\right) t\left(\tilde{\delta}\right)^{1-\varepsilon} 2\pi\tilde{\delta}d\tilde{\delta}dj \\ &= \left(\frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s}\right)^{1-\varepsilon} \int z_j^{\varepsilon-1} \kappa\left(n_{js}\right) dj \end{aligned}$$

Define $\mathcal{Z}_s^{\varepsilon-1} \equiv \left(\int z_j^{\varepsilon-1}\kappa\left(n_{js}\right)dj\right)$

$$P_s = \frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s\mathcal{Z}_s}$$

With this, we can simplify the expression for local profitability, x_s:

$$\begin{aligned} x_s &= \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^\varepsilon}\mathcal{L}_s c_s P_s^\varepsilon \left(B_s/W_s\right)^{\varepsilon-1} = \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^\varepsilon}\mathcal{L}_s P_s c_s \frac{\left(B_s/W_s\right)^{\varepsilon-1}}{P_s^{1-\varepsilon}} = \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^\varepsilon}\mathcal{L}_s P_s c_s \frac{\left(B_s/W_s\right)^{\varepsilon-1}}{\left(\frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s\mathcal{Z}_s}\right)^{1-\varepsilon}} \\ &= \frac{1}{\varepsilon}\mathcal{L}_s P_s c_s \left(\mathcal{Z}_s\right)^{1-\varepsilon} \\ &= \frac{1}{\varepsilon}\mathcal{L}_s \frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s\mathcal{Z}_s} c_s \left(\mathcal{Z}_s\right)^{1-\varepsilon} \\ &= \frac{1}{\varepsilon-1}\frac{W_s\mathcal{L}_s}{B_s\mathcal{Z}_s^\varepsilon} c_s \end{aligned}$$

A.3.2 Market clearing for Space

We use the same approach to characterize the total amount of local real estate used by plants. Consider a square of size $k \times k$. The fraction of of land devoted to commercial real estate is

$$\mathcal{N}_s^{k\Delta} = \xi\frac{1}{k^2}\int_{s\in S_i^k}\int 1\left\{j \text{ has plant in } s\right\} djds = \xi\frac{1}{k^2}\int N_{ji}\left(j\right)dj$$

where N_{ji} is the number of plants the firm places in square $\mathcal{S}_i^k$. Using $\xi = \Delta^2$, this is

$$\mathcal{N}_s^{k\Delta} = \Delta^2\frac{1}{k^2}\int N_i^k\left(j\right)dj$$

Using $n_{ji} = \frac{\Delta^2 N_{ji}}{k^2}$

$$\mathcal{N}_s^{k\Delta} = \int n_{ji}dj$$

Taking the limit as $\Delta \to 0$ and $k \to 0$ gives

$$\mathcal{N}_s = \int n_{js} dj$$

A.3.3 Consumption

We derive here an expression for the local consumption bundle. Labor used by firm j in a plant located in o to produce $c_{js}\mathcal{L}_s$ units of output for consumption by households in location s is

$$\begin{aligned} l_{jos}(\delta) &= \frac{T(\delta_{os})}{B_o Z_j} c_{js}\mathcal{L}_s = \frac{T(\delta_{os})}{B_o Z_j} c_s P_s^{\varepsilon} p_{js}^{-\varepsilon} \mathcal{L}_s = \frac{T(\delta_{os})}{B_s Z_j} c_s P_s^{\varepsilon} \left[\frac{\varepsilon}{\varepsilon - 1} \frac{W_o T(\delta_{os})}{B_s Z_j}\right]^{-\varepsilon} \mathcal{L}_s \\ &= \left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \frac{1}{W_o} \left(\frac{W_o T(\delta_{os})}{B_o Z_j}\right)^{1-\varepsilon} c_s P_s^{\varepsilon} \mathcal{L}_s \end{aligned}$$

We again use the approach of studying a $k \times k$ square in which economic activity is uniform. In such a square, firm j sets up N_{ji} plants, each with a catchment area that is a regular hexagon of size $1/N_{ji}$ (again, ignoring boundary issues, which disappear in the limit as $\Delta \to 0$. If the density of employment in the square is $\mathcal{L}_i^{k\Delta}$ and consumption per capita is $c_i^{k\Delta}$, then, per unit of space, total employment of firm in the square is then

$$\frac{1}{k^2} N_{ji} \int_0^{\psi\sqrt{k^2/N_{ji}}} \left[\left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \frac{1}{W_i^{k\Delta}} \left(\frac{W_i^{k\Delta} T(\delta)}{B_i^{k\Delta} Z_j}\right)^{1-\varepsilon} c_i^{k\Delta} \left(P_i^{k\Delta}\right)^{\varepsilon} \mathcal{L}_i^{k\Delta}\right] \varpi\left(\frac{\delta}{\psi\sqrt{k^2/N_{ji}}}\right) 2\pi\delta d\delta$$

Employment across all firms per unit of space is then

$$\mathcal{L}_i^{k\Delta} = \int \frac{1}{k^2} N_{ji} \int_0^{\psi\sqrt{k^2/N_{ji}}} \left[\left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \frac{1}{W_i^{k\Delta}} \left(\frac{W_i^{k\Delta} T(\delta)}{B_i^{k\Delta} Z_j}\right)^{1-\varepsilon} c_i^{k\Delta} \left(P_i^{k\Delta}\right)^{\varepsilon} \mathcal{L}_i^{k\Delta}\right] \varpi\left(\frac{\delta}{\psi\sqrt{k^2/N_{ji}}}\right) 2\pi\delta d\delta dj$$

Using the change of variables $\tilde{\delta} = \delta/\Delta$ and $n_{ji} = \frac{\Delta^2}{k^2} N_{ji}$, this is

$$\begin{aligned} \mathcal{L}_i^{k\Delta} &= \int n_{ji} \int_0^{\psi/\sqrt{n_{ji}}} \left[\left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \frac{1}{W_i^{k\Delta}} \left(\frac{W_i^{k\Delta} t(\tilde{\delta})}{B_i^{k\Delta} Z_j}\right)^{1-\varepsilon} c_i^{k\Delta} \left(P_i^{k\Delta}\right)^{\varepsilon} \mathcal{L}_i^{k\Delta}\right] \varpi\left(\frac{\tilde{\delta}}{\psi/\sqrt{n_{ji}}}\right) 2\pi\tilde{\delta} d\tilde{\delta} dj \\ &= \int \left[\left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \frac{1}{W_i^{k\Delta}} \left(\frac{W_i^{k\Delta}}{B_i^{k\Delta} Z_j}\right)^{1-\varepsilon} c_i^{k\Delta} \left(P_i^{k\Delta}\right)^{\varepsilon} \mathcal{L}_i^{k\Delta}\right] \kappa(n_{ji}) dj \end{aligned}$$

Taking the limit as $\Delta \to 0$ and $k \to 0$ gives

$$\begin{aligned}\mathcal{L}_s &= \int \left[\left(\frac{\varepsilon}{\varepsilon - 1} \right)^{-\varepsilon} \frac{1}{W_s} \left(\frac{W_s}{B_s Z_j} \right)^{1-\varepsilon} c_s \left(P_s \right)^{\varepsilon} \mathcal{L}_s \right] \kappa(n_{js}) dj \\ &= \left[\left(\frac{\varepsilon}{\varepsilon - 1} \right)^{-\varepsilon} \frac{1}{W_s} \left(\frac{W_s}{B_s} \right)^{1-\varepsilon} c_s \left(P_s \right)^{\varepsilon} \mathcal{L}_s \right] \mathcal{Z}_s^{\varepsilon - 1}\end{aligned}$$

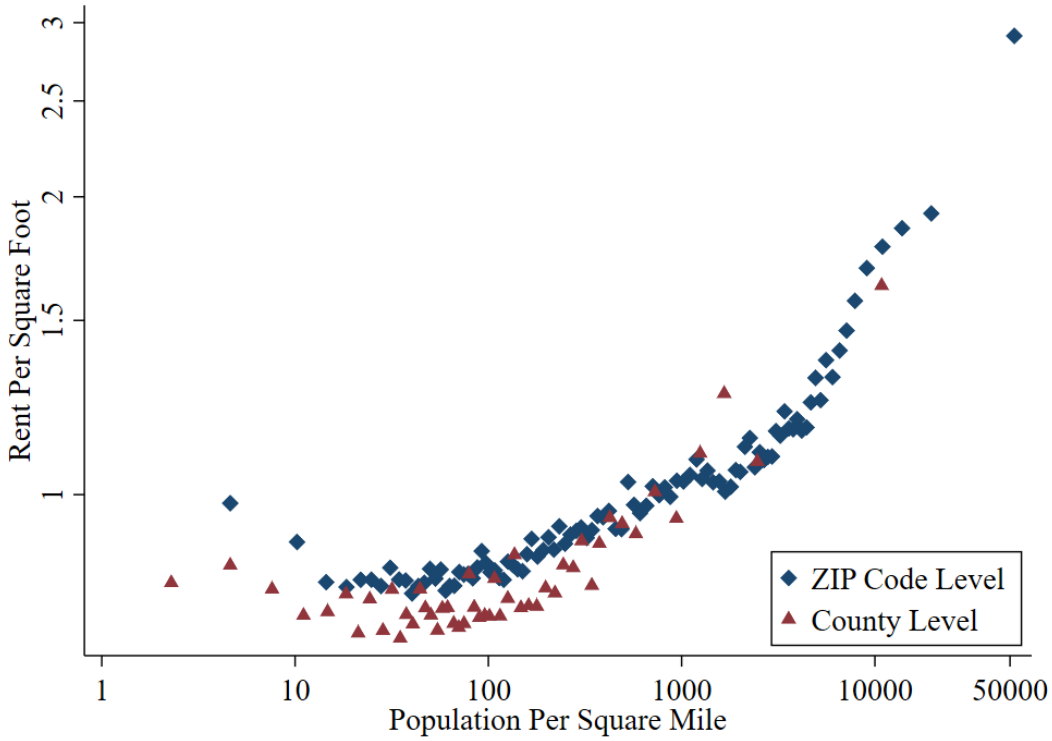
Combining this with the expression for the price level $P_s = \frac{\varepsilon}{\varepsilon - 1} \frac{W_s}{B_s \mathcal{Z}_s}$ and simplifying gives

$$c_s = B_s \mathcal{Z}_s$$

B Rents and Density

Our theory relies on the observation that more desirable locations exhibit a higher rent. As more desirable locations have higher population density, we can study empirically the connection between rents and a location desirability by testing whether locations with higher population density have higher rent. We borrow rent data from Zillow for the year 2018. For population density, we use the 2012-2016 population estimate provided in the American Community Survey (ACS) dataset. For each zipcode and county, Zillow provides an estimate of the rent per square foot. The rent per square foot is a preferable measure of rent than just the average rent in a location as the former controls for differences in housing size across locations, while the latter does not. Figure B.1 shows how the rent per square foot of a location, measured at either zip code or county levels, increases with the location population density.

Figure B.1: Rents across space



Notes: The figure presents the binned rent per square foot of a location (zipcode or county) in 2018 as a function of the location population density. Rent data comes from Zillow, while we use the 2012-2016 population estimate in the American Community Survey (ACS) to construct the population density measure.

C Sorting and Single-Crossing

In this section we perform some robustness checks to our sorting and single-crossing results.

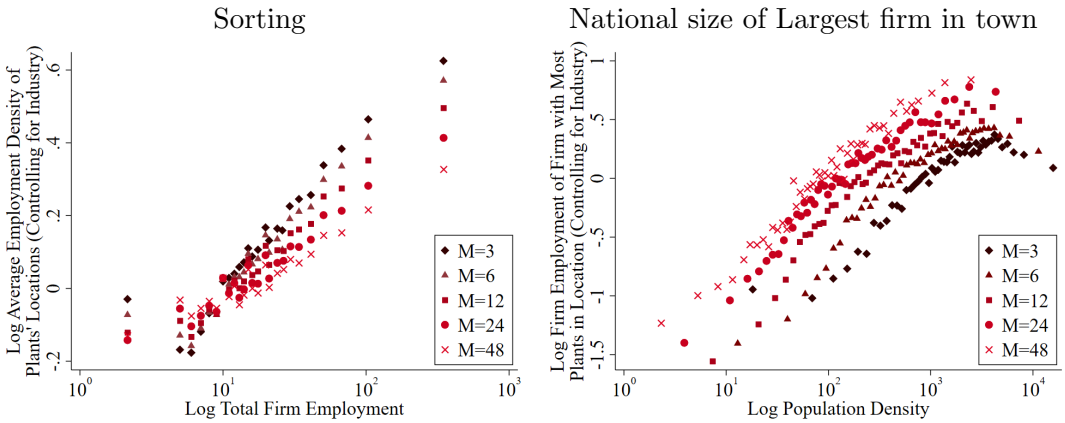
C.1 Excluding Each Firm's Own Employment

Figure C.1 revisits our sorting results both sorting and single-crossing results, but redefines the average population density (for sorting) and total firm employment (for single-crossing) to be net of the firm contribution to employment/population in a location (for sorting) and of the location contribution to total firm employment (for single-crossing).

C.2 Alternative Approaches to Identify the Largest Firm in Town

In Figure 15, in locations where multiple firms tie for the highest number of plants we use the average firm size among these firms. There are other reasonable approaches. We explore two alternative approaches in Figure C.2. In the left panel we exclude locations with ties for the highest number of plants. In the right

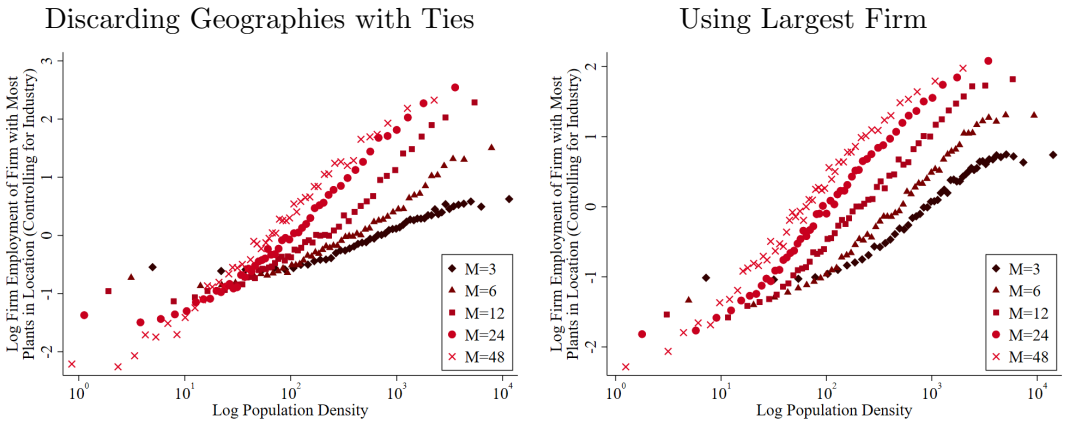
Figure C.1: Sorting, excluding own firm contribution to local employment



Notes: The left figure presents the results for the log average employment density of each location, weighted by the number of establishments of a particular firm operating in a particular industry in the location, net of the firm's employment contribution to each location's employment level. We subtract industry fixed effects, then bin the residuals by the log total firm employment at the national level. The right figure is produced by finding the log employment of the firm with the most plants in an industry and location, net of the location's contribution to the firm's total employment level, and regressing its total employment on industry fixed effects, weighted by each industry's total employment. In locations where multiple firms are tied for the highest number of plants, we take the average of the firm size. We then bin the residuals against log population density. Both figures present the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

panel, whenever there is a tie for the highest number of plants, we use the firm size of the largest firm among those tied. In both cases, as in Figure 15, we find patterns consistent with single-crossing.

Figure C.2: The National Size of the Largest Firm in Town, alternative methods

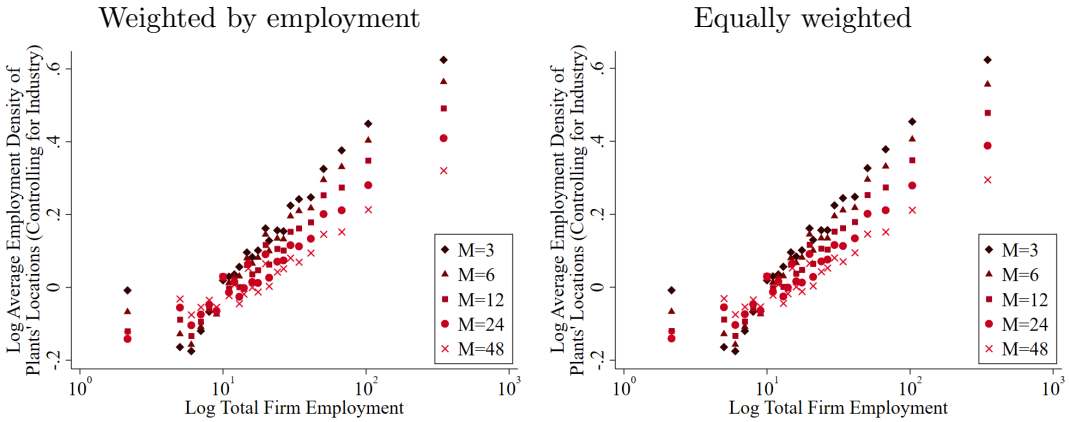


Notes: The figure is produced by finding the employment of the firm with the most plants in an industry and location, and regressing its total employment on industry fixed effects, weighted by each industry's total employment. In the left figure we discard locations where multiple firms are tied for the highest number of plants. In the figure on the right we use the employment of the largest firm in locations where multiple firms are tied for the highest number of plants. We then bin the residuals against log population density for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

C.3 Alternative Weights

Figure C.3 revisits the sorting results, but using alternative weighting methods. In the left plot, average population density is weighted by the number of plants of firm, while in the plot in the right average population density is equally weighted across locations where the firm is present.

Figure C.3: Sorting in the data, alternative weighting



Notes: The left figure presents the results where the average employment density of each location is weighted by the employment of a particular firm in the location. In the figure on the right each observation is equally weighted. To produce these figures we first subtract industry fixed effects, we then bin the residuals by the log total firm employment. Both figures present the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

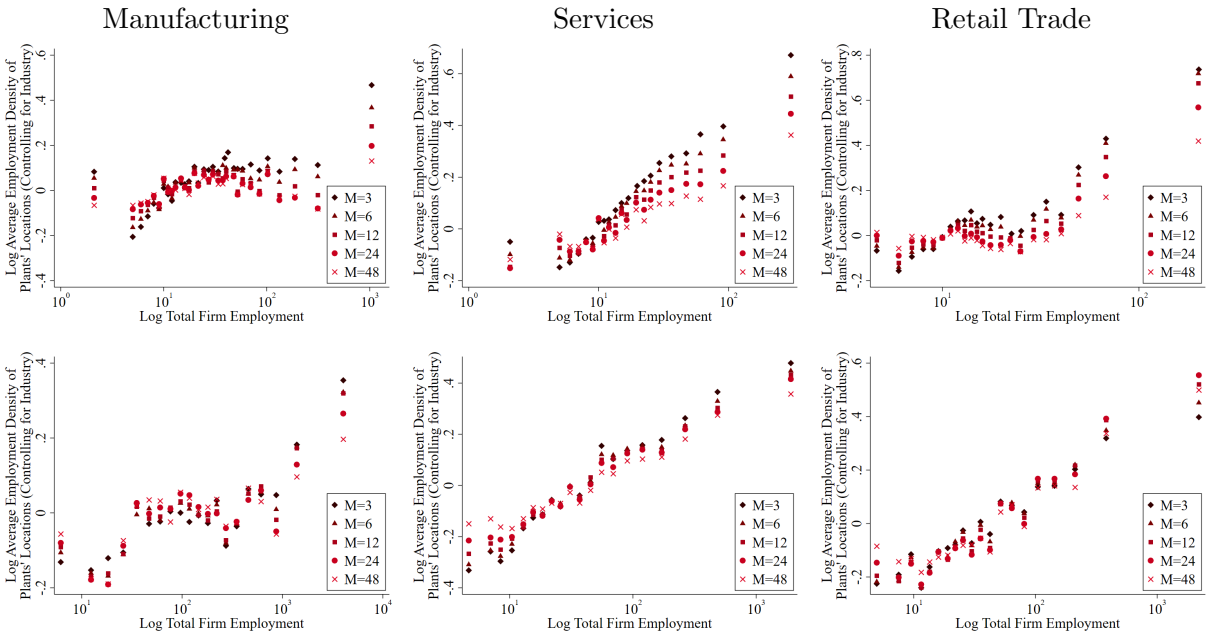
C.4 Heterogeneity in Sorting

In Figures C.4 and C.5 we study how sorting and the single-crossing property vary across major sectors. We focus on three major industries: manufacturing (left plots in both figures), services (center plots in both figures), and retail trade (right plots in both figures). The first row in each figure uses all firms within the industries, while plots in the second row exclude firms with a single plant in a given industry.

The first row of both Figures C.4 and C.5 paint a clear picture: Although less pronounced for manufacturing, sorting with the single-crossing property seems to be present in the three major sectors that we present. Not surprisingly, the effects seem to be weaker for manufacturing. Manufacturing has a higher prevalence of large firms with a single plant. For example, this could be due to lower transportation costs, or to higher returns to scale relative to other industries. Once we remove firms with a single plant within an industry, strong sorting patterns, exhibiting the single-crossing property, are clearly observed for the three major sectors. This can be seen in the second row of Figures C.4 and C.5.

There is clear evidence of sorting in all three sectors. However, it seems that the extent sorting varies by sector. The average density of a location where a firm locates its plants exhibits a flatter profile in manufacturing, a moderately increasing profile in services, and a strongly increasing profile for retail trade.

Figure C.4: Sorting by Sector



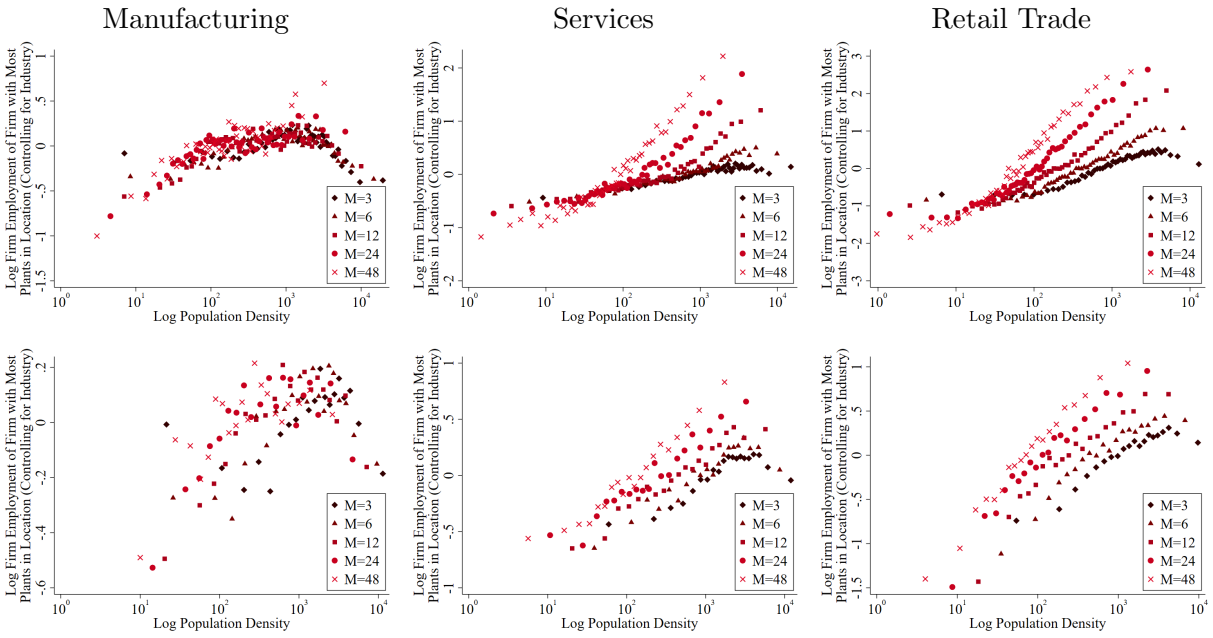
Notes: The figures present the results where the log average employment density of each location is weighted by the number of establishments of a particular firm operating in a particular industry in the location. To produce the figures, we subtract industry fixed effects, and bin the residuals by the log total firm employment at the national level. The figures present the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles. The left plots presents the results for the manufacturing sector, the center plots present the results for the service sector, while the right plots present the results for retail. The plots in the first row present the results using all firms within a sector. The plots on the second row remove firms with only one plant.

The same is true for the the size of the firm with most plants in a location as a function of the rent of that location. In other words, it seems that retail exhibits the highest amount of sorting across space, services exhibits moderate amount of sorting across space, and manufacturing the least amount of sorting across space.

C.5 Non-Imputed Data

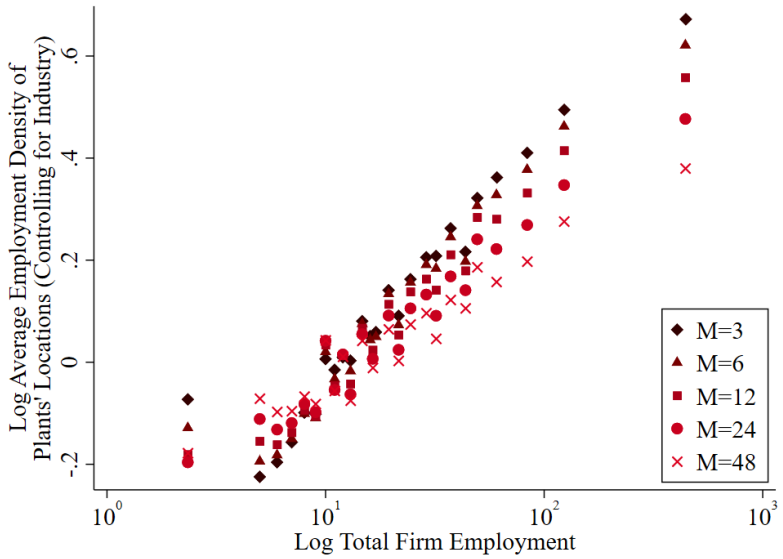
In this section we show that our sorting results are robust to using only non-imputed data. Figure C.6 shows that larger firms set plants in denser locations. Figure C.7 shows that firms that grow move towards denser locations. Finally, Figure C.8 shows that the size of the largest firm in town increases with the density of the location.

Figure C.5: The National Size of the Largest Firm in Town by Sector



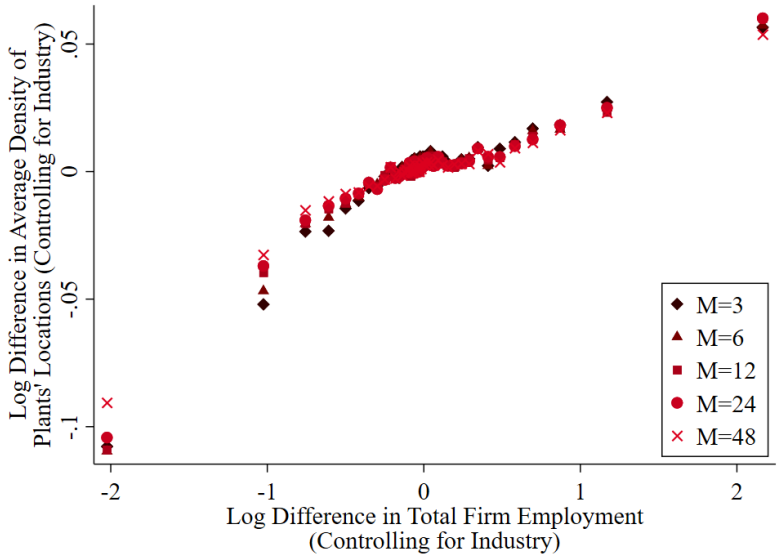
Notes: The figures are produced by finding the log employment of the firm with the most plants in an industry and location, and regressing its total employment on industry fixed effects, weighted by each industry's total employment. In locations where multiple firms are tied for the highest number of plants, we take the average of the firm size. We then bin the residuals against log population density for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles. The left plots present the results for the manufacturing sector, the center plots present the results for the service sector, while the right plots present the results for retail. The plots in the first row present the results using all firms within a sector. The plots on the second row removes firms with only one plant.

Figure C.6: Sorting in the data, non-imputed data



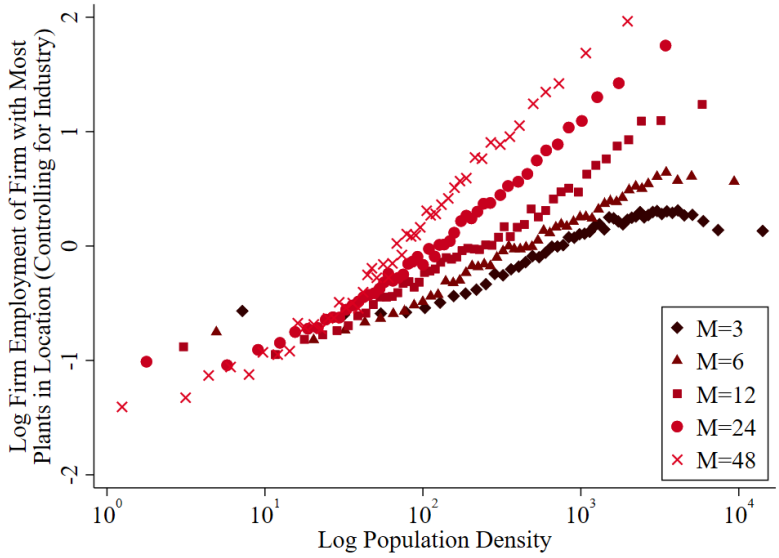
Notes: The figure presents the results for the log of the average employment density of each location, weighted by the number of establishments of a particular firm operating in a particular industry in the location. To produce the figure we first subtract industry fixed effects. We then bin the residuals by the log total firm employment at the national level. The figure presents the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

Figure C.7: Sorting over time, non-imputed data



Notes: The figure is produced by calculating the average population density across all locations where a firm is present, for a given industry, holding densities at their 2000 levels and weighting by the firm's establishments in the location in both 2000 and 2014, for firms-industry pairs with positive employment in both years. We then compute the log-difference in firm size and in average population density between 2000 and 2014, regress each of them separately on industry fixed effects, and store the residuals from each regression. We bin the residuals of the log-difference in average density by the log-difference in firm size residuals. The figure presents the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

Figure C.8: The national size of the largest firm in town, non-imputed data



Notes: The figure is produced by finding the log employment of the firm with the most plants in an industry and location, and regressing its total employment on industry fixed effects, weighted by each industry's total employment. In locations where multiple firms are tied for the highest number of plants, we take the average of the firm size. We then bin the residuals against log population density for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

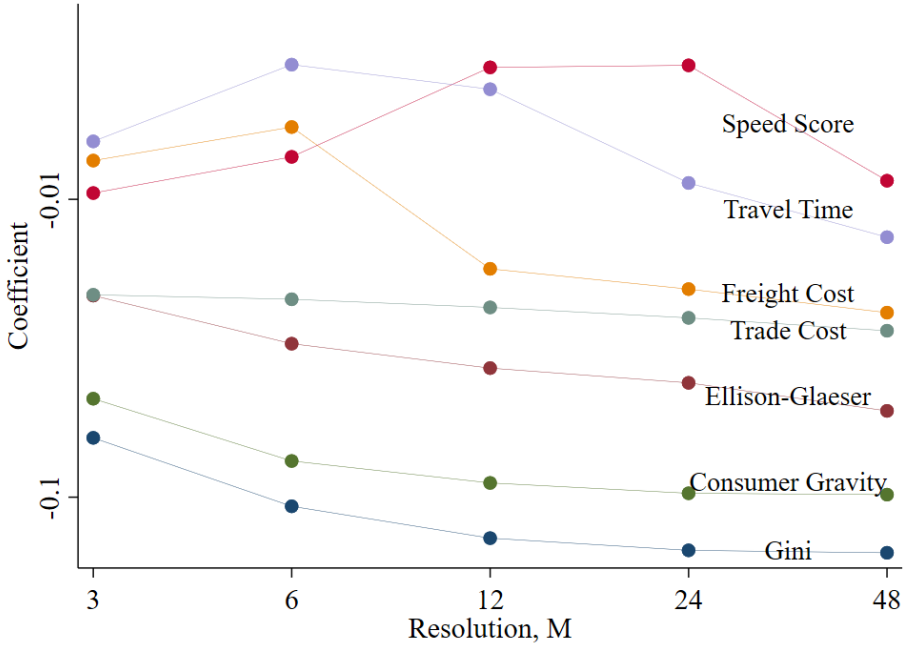
D Transportation Measures and Additional Results

In this section we provide a description of the transportation measures that we construct to asses the importance of transportation efficiency in explaining the way firms set up their stores across space (see Section 4.5). We build seven different transportation measures. Some of them vary across firms across industries, while other vary within a firm across space. A description of these measures is available in Table IV.

Table IV: Measures of Transportation Efficiency

Measure	Description
Gini coefficient	Gini coefficient of industry i's employment in a location relative to population in that location. A high value indicates that industry i is concentrated relative to population. We interpret this as industry i exhibiting high transportation efficiency. Constructed at SIC8 level.
Ellison-Glaeser	Constructed following Ellison and Glaeser (1997). Compares firm employment distribution in industry i with counterfactual distribution where firms locate randomly across space. A high value indicates high concentration. We interpret this as high transportation efficiency. Constructed at SIC8 level.
Consumer gravity	Constructed following Agarwal et al. (2017). The index measures, across industries, the elasticity of expenditure with respect to distance. We map industries with weaker gravity to industries with high transportation efficiency. This index is constructed for a subset of SIC2 industries. We apply the same index value to all SIC8 industries within an SIC2 industry with an estimate.
Freight Cost	Additive inverse of 1992 measure constructed in Bernard et al. (2006). Available for manufacturing industries at SIC2 level. We apply the same index value to all SIC8 industries within an SIC2 industry with an estimate.
Trade Cost	Additive inverse of measure constructed in Gervais and Jensen (2019). Trade cost estimates for 969 service and manufacturing industries at NAIC6 level. We apply the same index value to all SIC8 industries within a NAICS6 industry with an estimate.
Speed Score	Constructed following Couture et al. (2018). Provides Estimate of driving speed for the top 50 Metropolitan Statistical Areas (MSAs) in the US, for 2008. We assign the same speed score to all locations within an MSA.
Travel Time	Constructed using data for 2014 from the Texas Transportation Institute. Provides travel time at peak period relative to free-flow conditions for the major 470 urban areas in the US. We assign the same travel time to all locations within an urban area. We apply an additive inverse to the measure provided by the Institute.

Figure D.1: The Effect of Transportation Efficiency on Average Plant Employment Across Resolutions



Notes: The figure presents the estimated coefficients on the interaction term, γ_1, in equation 13 for each of the transportation efficiency measures described in Table IV. All transportation efficiency measures are standardized. We present the results for five different resolutions M: 3 miles, 6 miles, 12 miles, 24 miles, and 48 miles.

E Numerical Exploration: Algorithm

In this section we describe the algorithm that we used to solve for the industry equilibrium. Our algorithm exploits the first order conditions of the firm's problem (equations 3 and 4),

$$x_s z_j^{\varepsilon-1}\kappa'(n_{js}) \leq R_s + \lambda_j, \qquad \text{with equality if } n_{js} > 0, \text{ and}$$
$$\lambda_j = -\frac{d[z(q_j, N_j)^{\varepsilon-1}]}{dN_j}\int_s x_s \kappa(n_{js})ds,$$

where $z_j = z(q_j, N_j)$ with $N_j = \int_s n_{js}ds$, $x_s = \frac{I_s/(\varepsilon-1)}{\mathcal{Z}_s^{\varepsilon-1}}$ with $\mathcal{Z}_s = \left(\int_j z_j^{\varepsilon-1}\kappa(n_{js})dj\right)^{\frac{1}{\varepsilon-1}}$, and $R_s = R(I_s)$.

Our algorithm iterates on three univariate functions, $\mathcal{Z}_s \; \forall \; s$, and $\{N_j, \lambda_j\} \; \forall \; j$. Let $t = 0, 1, 2, ...$ denote the iteration round. Given an initial guess or the results of the previous iteration, $\mathcal{Z}_s^{t-1} \; \forall \; s$, and $\{N_j^{t-1}, \lambda_j^{t-1}\} \; \forall \; j$, we can compute the following objects: (i) $n_{j,s}^t \; \forall \; j, s$ (using equation 3), (ii) $N_j^t = \int_s n_{js}^t ds$, (iii) $z_j^t = z(q_j, N_j^t)$, (iv) $\mathcal{Z}_s^t = \left(\int_j \left(z_j^t\right)^{\varepsilon-1}\kappa(n_{js}^t)dj\right)^{\frac{1}{\varepsilon-1}}$, (v) $x_s^t = \frac{I_s}{(\varepsilon-1)(\mathcal{Z}_s^t)^{\varepsilon-1}}$, and (vi) $\lambda_j^t = -\int_s \frac{\partial z(q_j, N_j^t)^{\varepsilon-1}}{\partial N} x_s^t \kappa(n_{js}^t)ds$. We

repeat this procedure until a convergence criterion is satisfied, that is, until there is a $t = \tilde{t}$ such that

$$||\mathcal{Z}_s^{\tilde{t}} - \mathcal{Z}_s^{\tilde{t}-1}|| + ||N_j^{\tilde{t}} - N_j^{\tilde{t}-1}|| + ||\lambda_j^{\tilde{t}} - \lambda_j^{\tilde{t}-1}|| \leq \epsilon \ ,$$

where $||\cdot||$ is the sup norm and ϵ is a small number.

We use a two dimensional grid of points to numerically integrate when necessary and to evaluate the convergence criterion. Specifically, we use a two dimensional grid of S locations and J firms. For each iteration, a sufficient state is the value of the functions N_j, λ_j, and $\mathcal{Z}_s^t$, at these grid points. For each point j, s on the grid, we require only the values of $\mathcal{Z}_s^{t-1}$, N_j^{t-1}, and λ_j^{t-1} to evaluate n_{js}^t. To find N_j^t, numerically integrate across the locations using the trapezoid rule and the values of n_{js}^t at each of the S location grid points. Similarly, to find λ_j^{t-1}, we numerically integrate across locations. To find each location's local productivity, for any s, we numerically integrate across firms using the values of z_j^t and n_{js}^t at each grid point. This delivers new values of the functions at each of the S grid points. Finally, to evaluate the norms, we evaluate the convergence criterion by numerically integrating using the grid points.

In our numerical simulation, we used $J = 50$ and $S = 100$, but we found no noticeable difference in the solution when we used a grid of $J = 30$ and $S = 50$.

A complication arises due to the fact that $\kappa(n)$ is linear in the neighborhood of $n = 0$ (see Lemma 2). Because of this linearity, n_{js} move quite a bit across iterations in response to very small changes in $\mathcal{Z}_s$, N_j and λ_j. This can generate cycles in the iteration process. We handle this issue in two ways.

First, at each iteration, we do not fully update the policy functions. That is, we evaluate iteration $t+1$ using $\tilde{\mathcal{Z}}_s^t$, $\tilde{N}_j^t$ and $\tilde{\lambda}_j^t$ instead of $\mathcal{Z}_s^t$, N_j^t and λ_j^t, where $\tilde{\mathcal{Z}}_s^t = \varsigma \mathcal{Z}_s^{t-1} + (1-\varsigma)\mathcal{Z}_s^t$, $\tilde{N}_j^t = \varsigma N_j^{t-1} + (1-\varsigma)N_j^t$, and $\tilde{\lambda}_j^t = \varsigma \lambda_j^{t-1} + (1-\varsigma)\lambda_j^t$, where $\varsigma \in (0,1)$ is a dampening parameter. In principle, there exists a $\varsigma < 1$ such that cycles are not a concern. However, in many situations (i.e. sets of parameter values) the low degree of updating of the policy functions makes the code extremely slow.[30] Thus, we take an additional step.

Second, we replace the function $\kappa(n)$ with

$$\hat{\kappa}(n) = \alpha \mathcal{H}(n) + (1-\alpha)\kappa(n) \ ,$$

where $\mathcal{H}(n) = 1 - e^{-n/h}$. Notice that $\mathcal{H}'(n) > 0$, $\mathcal{H}''(n) < 0$, with $\mathcal{H}(0) = 0$, $\lim_{n\to\infty} \mathcal{H}(n) = 1$.[31] As a result, $\hat{\kappa}'(n) > 0$, $\hat{\kappa}''(n) < 0$, with $\hat{\kappa}(0) = 0$, $\lim_{n\to\infty} \hat{\kappa}(n) = 1$.

That is, in the iteration process we use $\hat{\kappa}(n)$ instead of $\kappa(n)$. In our experiments, a combination of $\varsigma > 0$ and $\alpha > 0$ are able to handle cycles and thus allows the code to converge quickly. For the numerical explorations presented in this paper we use $\varsigma = 0.97$ and $\alpha = 0.0001$.

To ensure that this approximation yields an accurate solution, we can evaluate whether the resulting policy function found using $\hat{\kappa}$ is the solution to each firm's true problem that uses κ. Let $\hat{\mathcal{Z}}_s$, $\hat{N}_j$ and $\hat{\lambda}_j$ denote the solution of the iteration process (i.e. once the convergence criterion is satisfied) when we solve

[30] For high enough values of ς the code can take many hours to converge, even when J and S are small.

[31] The parameter h allows us to modify the concavity of the function $\mathcal{H}(n)$. Here, we used $h = 0.01$.

the firms problem using $\hat{\kappa}(n)$. We can also easily obtain $\hat{n}_{js}$, $\hat{z}_j = z(q_j, \hat{N}_j)$, and $\hat{x}_s$. To gauge the accuracy of the approximate solution, we compute,

$$\text{absolute error} = \int_s \int_j \mathbf{1}\left[\hat{n}_{js} > 0\right] \overbrace{\left[\hat{x}_s \hat{z}_j^{\varepsilon-1} \kappa'(\hat{n}_{js}) - (R_s + \hat{\lambda}_j)\right]}^{\text{error in equation 3}} djds \ ,$$

$$\text{relative error} = \frac{\int_s \int_j \mathbf{1}\left[\hat{n}_{js} > 0\right] \left[\hat{x}_s \hat{z}_j^{\varepsilon-1} \kappa'(\hat{n}_{js}) - (R_s + \hat{\lambda}_j)\right] djds}{\int_s \int_j \mathbf{1}\left[\hat{n}_{js} > 0\right] (R_s + \hat{\lambda}_j) djds} \ .$$

That is, the first expression computes the absolute error of the allocation using $\hat{\kappa}(n)$, but evaluating the first order condition using $\kappa(n)$, while the second expression provides the absolute error, relative to the level of costs for firm j in location s, as described by the RHS of equation 3. For our baseline equilibrium, we find that absolute error $= 0.00008$, and relative error $= 0.000025$. That is, the absolute error is 0.0025% of the average level of the RHS of the first order condition. This provides reassurance that the solution under $\hat{\kappa}(n)$ is a good approximation of the actual solution.