Plants in Space*

Ezra Oberfield
Princeton University

Esteban Rossi-Hansberg
University of Chicago

Pierre-Daniel Sarte
Federal Reserve Bank of Richmond

Nicholas Trachter
Federal Reserve Bank of Richmond

February 25, 2023

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 predictions on sorting across space for industries with many plants per firm. Compared with less productive firms, productive firms place more plants in dense high-rent locations 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. We present evidence consistent with these and several other predictions using U.S. establishment-level data.

*Oberfield: edo@princeton.edu. Rossi-Hansberg: rossihansberg@uchicago.edu. Sarte: pdgs4frbr@gmail.com. Trachter: trachter@gmail.com. We thank Milena Almagro, Fernando Alvarez, Donald Davis, Felix Tintelnot, Harald Uhlig, Conor Walsh, as well as participants at numerous seminars and conferences for their feedback. We thank Eric LaRose, Reiko Laski, James Lee, Samira Gholami, 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

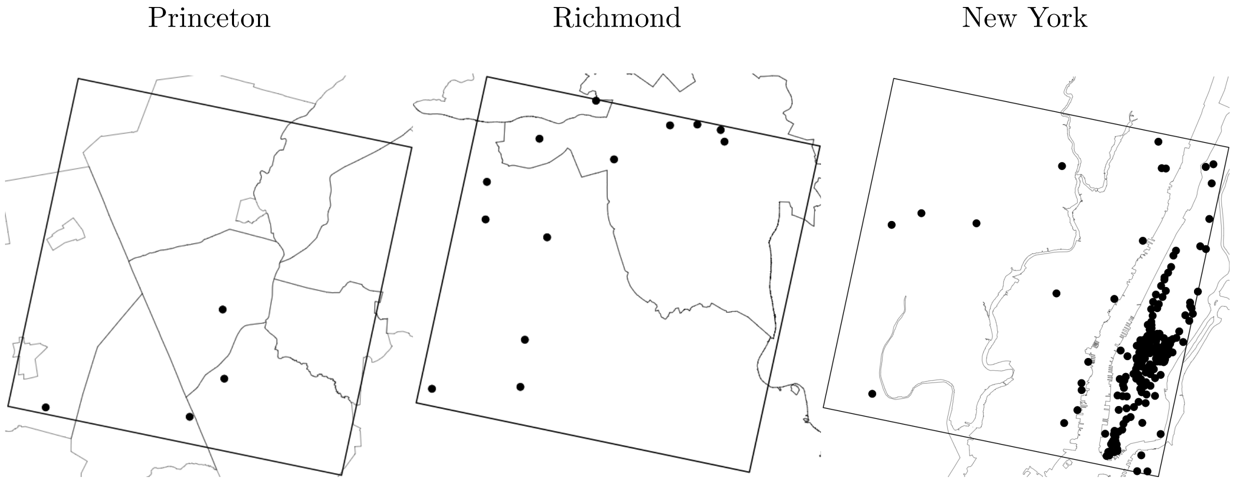
Delivering products and services to locations where consumers can easily access them involves complex decisions about where to locate plants and how large these plants should be. Having too few establishments is costly because it increases the distance to consumers. Having too many involves large span-of-control 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 around 14,000 stores in 2019 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 is 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. 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 has more stores than Walgreens in less dense locations. Is this form of sorting across locations a general feature of the solution to the location problem or of the data?

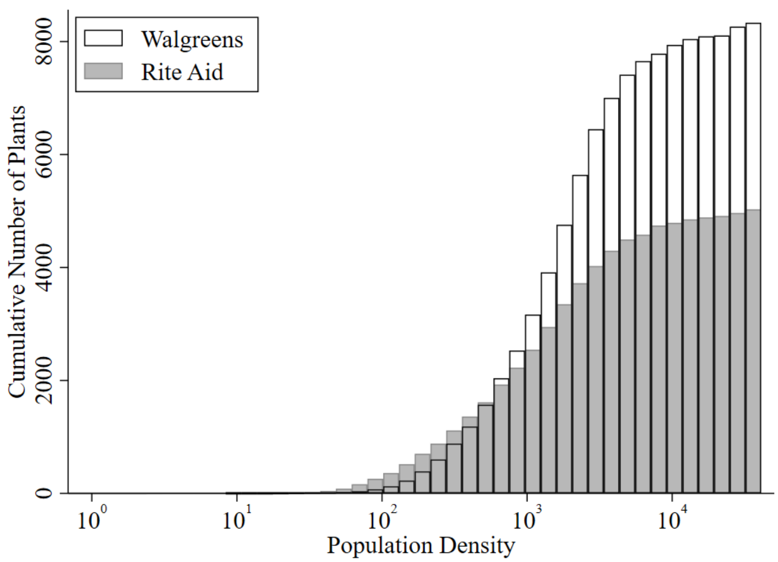
More generally, we aim to provide insights into two main questions: First, which firms set up plants in which locations? Second, what determines the scale and location of production? Answering these questions requires us to think about plants and firms as distinct, albeit related, economic entities. 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 demand for a firm's product. We focus primarily on the problem of a firm that takes these distributions as

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, using data from the 2010 decennial census taken from Manson et al. (2021).

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 its 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, exploiting 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. When space is one-dimensional, plants should be located at the center of equal-length intervals. In two dimensions, the result 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 its center to its customers. However, unlike hexagons, circles cannot be used repeatedly to form a tessellation. We extend the theorem to an environment where economic activity is heterogeneous across space. Specifically, 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 for industries that are well approximated by our limiting economy—industries for which catchment areas are small. First, more productive firms set up *more* plants in denser, high rent, locations than less productive firms. Perhaps more surprising is that they also set up *fewer* plants in markets

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

with lower density and rents. For highly productive firms, the span-of-control cost of managing additional plants in low-density locations outweighs the benefits of low rents. In contrast, for less productive firms, low local rents are more attractive since they manage fewer plants and so span-of-control costs are less relevant. Second, among firms that operate the same number of plants in a location, the plants from the more productive firm are larger in that location. Highly productive firms obtain large variable profits from each plant but limit the number of plants in a location due to the span-of-control cost they impose on the whole firm. Hence, when they set up the same number of plants as a less productive firm in a location, they choose to make these plants larger. 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. To illustrate this without adding too much additional structure, we embed the firm's problem into a small industry that does not affect local characteristics.[2] We show that, in equilibrium, the local industry price index is a function of local productivity and a weighted sum of the productivity of all firms present in the location, with weights that depend on each firm's local footprint which determines the cost of delivering goods to customers. These are the only local characteristics that determine equilibrium outcomes. Analyzing realistic quantitative 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. (2021) 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 (2022) 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 firm 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.

[2]Setting up the model in full general equilibrium with many industries and labor mobility is straightforward but requires making assumptions on a number of economic fundamentals that are unrelated to our main focus.

Ramondo and Rodríguez-Clare (2013) and Arkolakis et al. (2018) use a probabilistic structure that is remarkably tractable. Each plant's technology is constant returns to scale, so decisions on how to serve each market are independent across markets. Introducing fixed costs of setting up plants or span-of-control costs would render the model intractable. In general, the firm's plant location problem can be split into two parts: an inner problem of establishing the set of locations serviced by each of a firm's plants, their "catchment areas", and an outer problem of determining the location and number of plants. Tintelnot (2016) introduces fixed costs and solves the inner problem by assuming that firms sell a continuum of products and that plant productivities follow an extreme value distribution, which implies that plants sell to all locations. This smooths out the firm's objective function but still requires solving the combinatorial problem of where to set up plants and how many. All frameworks in this literature either solve the combinatorial problem with only a few countries or assume away fixed costs of setting up new plants. Methodologically, our main contribution is to solve the outer problem and characterize it analytically in a limiting case where all costs remain relevant. In contrast to the multinational literature, which has focused mostly on manufacturing, the particular limiting economy we study is likely to be a better approximation to industries with high transport costs, where the number of plants per firm is large. Substantially, our main contribution is to incorporate span-of-control costs to the firm's problem and to characterize the resulting sorting patterns. Most models in the literature 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.[3] 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 new plants reduces the value of other plants, cannot easily be solved except for algorithms that, at worst, evaluate all possible combinations. Recently, Arkolakis et al. (2017) and Hu and Shi (2019) have developed algorithms to solve these types of "submodular" problems more efficiently by iteratively pruning the choice set, 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.

There is a large, active literature in Operations Research studying the facility location problem. The classic Weber problem of placing a single plant to serve many destinations at minimal cost (Weber (1909)) was generalized to study the placement of multiple plants by Stollsteimer (1961) and Balinski (1965). There are many versions of the problem. One approach, used by much of the economics literature, studies the problem with a finite set of discrete locations. The alternative, which we follow, models space as continuous. Another distinction is whether or not there are limits on each plant's output (these are known

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

as the capacitated or uncapacitated facility location problems). Given the complexity of the problem, the literature has focused on numerical algorithms that deliver approximate solutions in polynomial time.[4]

We provide a characterization of the matching of heterogeneous firms with multiple plants to heterogeneous locations. Nocke (2006), 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 (2021), 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.[5]

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 an industry equilibrium. 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 all 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, $\mathcal{S} = [0,1]^{\mathtt{d}} \subset \mathbb{R}^{\mathtt{d}}$, where $\mathtt{d} \in \{1,2\}$ is the dimension of the space.[6] 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 its productivity, q_j. It chooses a finite set of locations $O_j \subset \mathcal{S}$ where to set up plants. Conditional on operating a plant at location o, production requires only local labor which is employed at wage W_o. A plant's productivity is the product of a local component, B_o, and a firm component, $Z(q_j, N_j)$. The firm component is increasing in a standard idiosyncratic productivity level q_j, and decreasing in the firm's total number of plants $N_j = |O_j|$. The latter captures the productivity cost of increasing the firm's span of control. We also assume that $Z(q, 0) < \infty$. In sum, if a firm operates a total of N_j plants, its productivity

[4]These problems have been shown to be NP hard in both one or two dimensions (Fowler et al. (1981)). It has been shown that, unless $P = NP$, there is a bound on the performance of such algorithms: they cannot guarantee a solution that is better than 1.463 times the actual minimal cost (Guha and Khuller (1999) and Sviridenko (2002)). Algorithms that deliver performance close to this bound have been recently proposed by Byrka and Aardal (2010) and Li (2013).

[5]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.

[6]Our results can be easily generalized to a Euclidean space $\mathcal{S}$ that is closed, bounded, and Jordan measurable. While it seems intuitive that our results could be extended beyond two-dimensional space, doing so would require a resolution of the Gersho conjecture (Gersho (1979)), which remains open for three or more dimensions.

in location $o \in O_j$ is $B_o Z(q_j, N_j)$. 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 at distance δ, $T(\delta) \geq 1$ units must be shipped. We normalize $T(0) = 1$ and assume that $T(\delta)$ is strictly increasing, satisfies the triangle inequality, and diverges as $\delta \to \infty$.

We posit a market structure in which firms transport goods to households and choose a separate price for its good at each destination.[7] 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 $\frac{W_o T(\delta_{so})}{B_o Z(q_j,N_j)}$ where δ_{so} denotes distance between s and o. 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. Let p_{js} be the price charged by j to consumers in s. Then, if $D_s(p_{js})$ is 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 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, $p_{js} = \frac{\varepsilon}{\varepsilon-1}\Lambda_{js}$.

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

$$\pi_j = \max_{O_j} \left\{ Z\left(q_j, N_j\right)^{\varepsilon-1} \int_s D_s \max_{o \in O_j} \left\{ b_o T\left(\delta_{so}\right)^{1-\varepsilon} \right\} ds - \sum_{o \in O_j} R_o \xi \right\}. \tag{2}$$

where $b_o \equiv \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}} \left(B_o/W_o\right)^{\varepsilon-1}$ summarizes local productivity and wages.

[7] An alternative market structure—one in which households incur shipping costs and choose which plant to purchase from, and each firm chooses a separate price at each of its plants—is equally natural. If households have isoelastic residual demand, as we assume below, the two market structures yield the same revenue and employment for each plant and the same consumption and expenditures for each consumer. We focus on the market structure in which firms incur shipping costs because it is simpler to state and work with.

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 sells goods from the plant in o. Formally, let $CT(o)$ denote the catchment area of a plant in location o,

$$CT(o) = \left\{ s \in \mathcal{S} \text{ for which } o = \arg\max_{\tilde{o} \in O_j} \left\{ b_{\tilde{o}} T\left(\delta_{s\tilde{o}}\right)^{1-\varepsilon} \right\} \right\} \ . \tag{3}$$

$CT(o)$ can be empty if a plant's cost, relative to other nearby plants, is high enough.

Note 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 plants are placed. When space is one dimensional, $\mathtt{d} = 1$, and transport costs rise sufficiently fast with distance, catchment areas are simply a set of non-overlapping intervals covering $\mathcal{S}$, such that the cost of servicing costumers at the boundary is identical for both adjacent plants. Therefore, the size of the catchment area of a particular plant is decreasing in the plant's cost and increasing in the cost of adjacent plants. When transport costs do not rise fast with distance, the catchment area of a plant can be the union of disjoint segments. With 2-dimensional space, catchment areas can be substantially more complicated. In any of these cases, it is straightforward, although computationally costly, to solve numerically for catchment areas.[8] Nevertheless, we show that when we incorporate the decision of *where* to place plants, the optimal choice leads to a structure of catchment areas that greatly simplifies the problem. In particular, in the limiting case that we study, local catchment areas are always characterized by intervals ($\mathtt{d} = 1$) or hexagons ($\mathtt{d} = 2$).

2.1.1 Examples in one and two dimensions

This section illustrates the role of the production cost in determining catchment areas. We first focus on the simpler case of $\mathtt{d} = 1$ with $\mathcal{S} = [0, 1]$, so that catchment areas partition the unit interval. We show that non-convex catchment areas can arise when transport costs do not rise quickly with distance. We then turn to an environment with two dimensions where $\mathcal{S}$ is the unit square. We show examples of how changes in fundamentals across locations affect catchment areas. For each exercise, we set $\varepsilon = 2$ and solve two cases with different distributions for b_o across space.[9]

Figure 3 presents the one-dimensional case. We consider the problem of a firm that (perhaps sub-optimally) places five plants at regular intervals across space.[10] We assume transport costs are given by $T(\delta_{so}) = 1 + 0.75\delta_{so}$, where δ_{so} is the Euclidean distance between s and o. The left panel presents the resulting catchment areas when production costs are the same across all locations, i.e. $b_o = 1 \ \forall \ o \in O_j$, while the right panel shows the resulting catchment areas when the third location is 14.5% more productive than

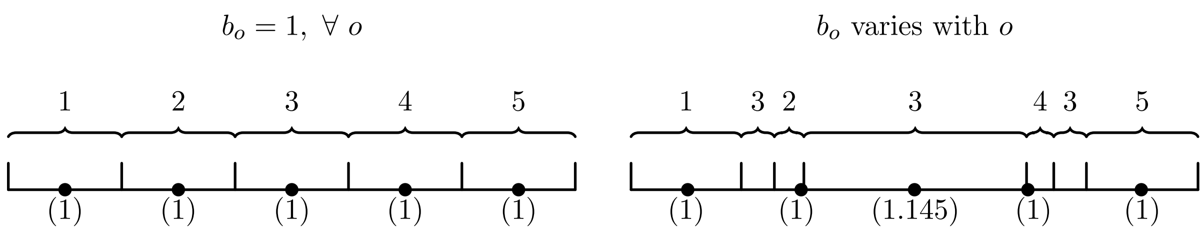
[8] This problem is equivalent to constructing a weighted Voronoi diagram. Tintelnot (2016) approaches this problem by assuming that plants sell a continuum of goods produced with productivities drawn from a distribution with infinite support. The implication is that the catchment areas of all plants overlap and cover the whole space.

[9] The choice $\varepsilon = 2$ is not essential to generate these examples.

[10] Plants locations are $o \in \{1/10,\ 3/10,\ 5/10, 7/10,\ 9/10\}$.

in the left panel. In both cases, catchment areas are characterized by a collection of segments. When all plants face the same cost, catchment areas are equally sized line segments. When one plant, in this example Plant 3, faces a lower cost, catchment areas vary in size and need not be convex. More productive locations have larger catchment areas, as evidenced by the spatial expansion of Plant 3's catchment area. Notice that Plant 3's catchment area depends on its own cost, its location, and the cost of its neighbors—Plants 2 and 4, as well as the location and cost of its neighbors' neighbors—Plants 1 and 5. In other words, designing the catchment area of plant 3 requires understanding this plant's interactions with all other plants.[11] Note also that the non-convexity in the catchment area of Plant 3 is possible because transport costs do not rise grow quickly with distance.

Figure 3: Catchment Areas in One Dimension



Notes: The figure presents the catchment areas, as defined in (3), for the one-dimensional case where 5 plants are located in a line. The left plot presents the case where all locations are equally productive, while the right plot presents a case where plant 3's location is more productive. Each dot in each plot corresponds to the location of a plant. Each number in parenthesis under each dot corresponds to the value for b_o for that dot. The number above each brace indicates which plant serves that location, i.e. the catchment area of a particular plant.

Similar logic applies for the two-dimensional case. Figure 4 presents catchment areas when nine plants are arbitrarily placed in a regular grid.[12] For these exercises we set transportation costs equal to $T(\delta_{so}) = 1+\delta_{so}$. As in the one-dimensional case, we assume that there is no variation in economic fundamentals, so $b_o = 1\ \forall\ o \in O_j$, in the left panel. The right panel presents the results when we increase the costs of the location in the upper left corner by setting productivity to $b_o = 0.85$, and we reduce the costs of the central and lower right corner locations by letting $b_o = 1.2$.

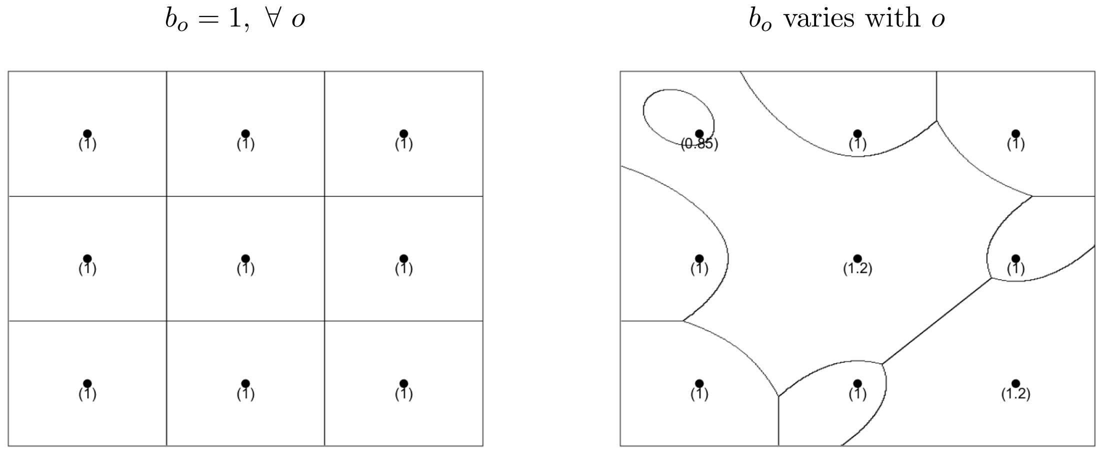
As Figure 4 shows, when production costs are constant across production locations, catchment areas are all equally sized, are all convex, and are all polygons. However, when production costs vary, the catchment areas can take different shapes and sizes, can be non-convex, and are not polygons.

While the one-dimensional case is simpler in the sense that catchment areas are, at most, a combination of disjoint intervals, both the one-dimension and two-dimension cases share the same basic features and

[11]One can show that if production costs are similar across space and if trade costs rise sufficiently fast with distance, a plant's catchment area depends only on its cost and that of its direct neighbors.

[12]In the figure, plants are located at $(1/6, 1/6), (1/6, 1/2), (1/6, 5/6), (1/2, 1/6), (1/2, 1/2), (1/2, 5/6), (5/6, 1/6), (5/6, 1/2), (5/6, 5/6)$.

Figure 4: Catchment Areas in Two Dimensions



Notes: The figure present the catchment areas, as defined in (3), 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 for that dot.

complexities: local characteristics affect the size of catchment areas, and non-convex catchment areas can arise as a combination of transport cost varying slowly with distance and specifics about the productivity of a location, the productivity and location of its neighbors, the productivity and location of the neighbors' neighbors, and so on.

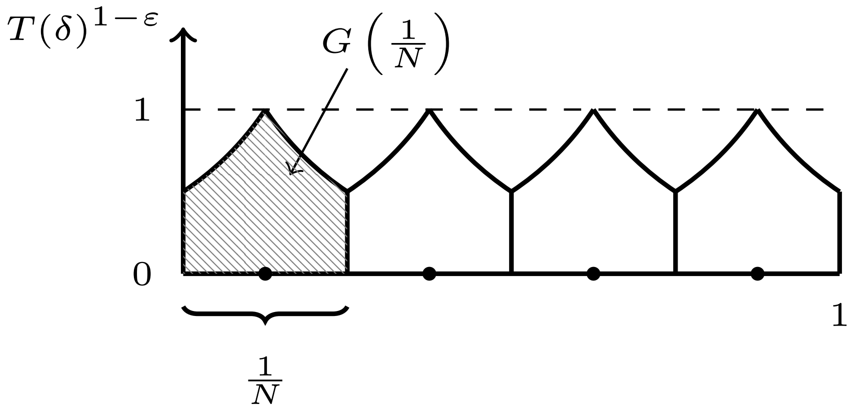
The difficulty of these problems may seem daunting, particularly once we introduce more locations, richer heterogeneity, and especially when we incorporate the outer problem of how many plants to use and where to place them. Perhaps surprisingly, in the limit economy we study below, solving this outer problem of optimal plant locations imposes structure that leads to a simple characterization of the inner problem of solving for catchment areas in both one and two dimensions. The only case where we can characterize the solution to the full problem without relying on the limit economy is the one-dimensional case with uniform locations. We turn to that problem first.

2.2 A Simple Special Case in One Dimension

We now turn to the outer maximization problem of determining how many and where to set up plants. We start by considering one special case for which the solution to this outer problem is straightforward. If space is one-dimensional and economic activity is uniform across locations, the optimal configuration is to have plants equally spaced so that catchment areas are equal in length. In this case, the firm's profit function can be expressed as

$$\pi_j = \sup_N xZ\left(q_j, N\right)^{\varepsilon-1} NG\left(1/N\right) - RN$$

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



where $x \equiv \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}} D\,(B/W)^{\varepsilon-1}$ and $G(u) = \int_{-u/2}^{u/2} T(|\delta|)^{1-\varepsilon} d\delta$. The variable x combines the demand, D, facing the plant at each location with the cost of effective labor, W/B, into a measure of local profitability. $G(u)$ is the integral of the function $T\,(\cdot)^{1-\varepsilon}$ over all distances between the origin and points of a line segment of length u centered at the origin.

We refer to the function $\kappa(N) \equiv NG(1/N)$ as the *efficiency of distribution*. 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 plants. Figure 5 provides a graphical representation of the function $G(\cdot)$ and the implied function $\kappa(\cdot)$. The figure presents an example with four plants, $N = 4$, with catchment areas of length $1/N$. It plots the function $T(\delta)^{1-\varepsilon}$, where δ is the distance from the plant to the customer's location. For customers at the plant's location, $\delta = 0$, there is no loss from transport costs. For customers further away, profits are reduced by a factor of $T(\delta)^{1-\varepsilon}$. The shaded area is $G\,(1/N)$, the fraction of profit the firm gets from a plant's catchment area relative to a world with no transport costs. $\kappa(N) = NG\,(1/N)$, the efficiency of distribution, is the fraction of profit the firm gets per unit of space, relative to a world with no transport costs.

In two dimensions, there is no closed-form solution for the exact placement of plants, even when economic activity is uniform across space. Nevertheless, there is a known upper bound for the profit a firm can attain, also based on the strategy of placing plants in a regular pattern across space. We discuss this strategy further in Section 2.4.

As discussed above, we aim to venture beyond these special cases in which economic activity is uniform, so that we can discuss how a firm's local footprint varies with local economic conditions, or how firms sort across space. Thus, we now propose a reformulation of this problem that can be tractably studied, while still preserving its main features and trade-offs.

2.3 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

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

for $\mathtt{d} = 1, 2$. 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. In two-dimensional space, $\mathtt{d} = 2$, ξ^{Δ} and Z^{Δ} depend on the square of Δ since space is two-dimensional while, in contrast, transport costs are a function of distance which is one dimensional.[13] The following proposition describes the profits of the firm in this 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 $x_s \equiv \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}} D_s \left(B_s/W_s\right)^{\varepsilon-1}$. In one dimension, $g(u)$ is the integral of the function $t(\cdot)^{1-\varepsilon}$ over all distances between the origin and points of a line segment of length u centered at the origin. In two dimensions, $g(u)$ is the integral of the function $t(\cdot)^{1-\varepsilon}$ over all distances between the origin and points of a regular hexagon of area u centered at the origin.[14]

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

[14] In one dimension, $g(u) = \int_{-u/2}^{u/2} t(|\delta|)^{1-\varepsilon} d\delta$. In two dimensions, $g(u) = \int_0^{\sqrt{3^{-3/2}2u}} t(\delta)^{1-\varepsilon} 2\pi\delta\varpi\left(\frac{\delta}{\sqrt{3^{-3/2}2u}}\right) d\delta$ where $\varpi(r)$ is the fraction of circle with radius r that intersects with the interior a regular hexagon with side length 1. As we show in Appendix A.1.2, simple trigonometric arguments yield $\varpi(\delta) = \begin{cases} 1 & 0 \le \delta \le \sqrt{3}/2 \\ 1 - \frac{6}{\pi}\cos^{-1}\left(\frac{\sqrt{3}/2}{\delta}\right) & \sqrt{3}/2 \le \delta \le 1 \end{cases}$.

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. As before, 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. Hence, in the limit, a location's characteristics can be fully summarized by two variables, local rent, R_s, and local profitability, x_s. In contrast, before taking the limit, the relevant features of a location were infinite dimensional, comprised of the local effective wage and demand in surrounding locations. Note also that, in the limit, aside from the span-of-control considerations, the problem is completely separable across locations. An important implication of this last result is that the problems in one- and two dimensions end up being incredibly similar. The only difference is that when solving the problem in one dimension the function $g(u)$ requires integrating over a line segment, while in two dimensions it requires integrating over a hexagon.

Before presenting a sketch of the proof of Proposition 1, we go back to the simpler case where all locations are identical that we analyzed in Section 2.2 for one-dimensional space. We then proceed to sketch the proof of Proposition 1 and characterize the solution to the profit maximization problem. Most formal proofs are relegated to the Appendix unless explicitly stated.

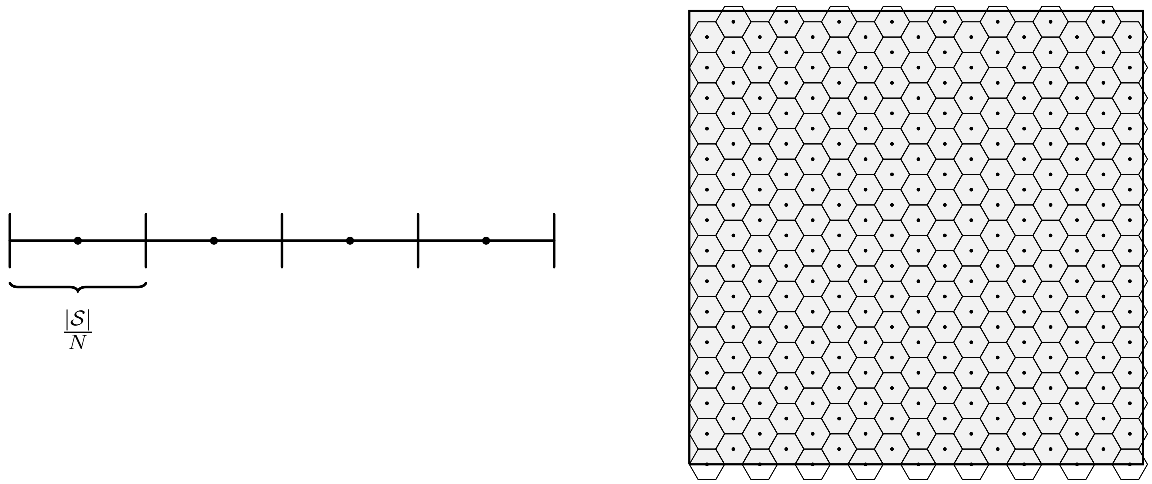
2.4 Uniform Space

We begin by discussing the simpler case where all economic activity is uniform across locations. We analyzed the one-dimensional version of this problem in Section 2.2. Here we also discuss the case of two dimensions where there is no closed-form solution for the placement of plants. As before, assume the local demand shifter for all locations is D, effective productivity is b, and commercial land rents are R. When space is homogeneous, there are some known results to the solution to the firm problem of choosing where to locate its plants. In particular, if a firm places N plants, the firm's payoff will be no higher than $xz(q_j, N)^{\varepsilon-1}NG\left(|\mathcal{S}|/N\right) - RN$ where $|\mathcal{S}|$ is the length of $\mathcal{S}$ in one dimension, and the area of $\mathcal{S}$ in two dimensions and, as stated in Proposition 1, $x \equiv Db$ and $G(u)$ is the integral of the transportation cost $T(\cdot)^{1-\varepsilon}$ over either a line segment of length u or a regular hexagon with area u centered at the origin, in each respective case. For the one-dimensional case, the optimality of catchment areas that are congruent line segments is trivial, as shown in the left panel of Figure 6: given all the symmetry built into the model, it is straightforward to set the catchment area of each plant to be $|\mathcal{S}|/N$, with plants placed at the center of each catchment area. In this case, as we argued in Section 2.2, $xz(q_j, N)^{\varepsilon-1}NG\left(|\mathcal{S}|/N\right) - RN$ is not only an upper bound but it is equal to the maximum firm's payoff when placing N plants in a one-dimensional uniform space.

When space is two-dimensional, the result follows from the Sum of Moments theorem in Fejes Toth (1953), one of the landmark results in discrete geometry.[15] 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. The right panel of Figure 6 shows an example of this solution. Why are hexagonal catchment areas optimal in two-

[15]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 6: Filling Out Space



Notes: The left panel in the figure shows a line of length $\mathcal{S}$ divided by $N = 4$ line segments of length $\mathcal{S}/N$. The right panel shows a square of area $|\mathcal{S}|$ divided by N hexagons of area $|\mathcal{S}|/N$.

dimensional space? 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. Note that this is an upper bound. As the right panel in Figure 6 shows, N disjoint uniform hexagons of size $|\mathcal{S}|/N$ generically do not fit exactly in the space $\mathcal{S}$.[16] 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 issue is quantitatively less relevant. In the "appropriate" limit, the upper bound is attained.

2.5 Heterogeneous Space

We are interested in understanding the location of a firm's plants in *heterogeneous* space. Section 2.4 provided important tools that we take advantage of for the general case with spatial heterogeneity. For the homogeneous space case, we know how to construct the solution to the firm's problem for the $\mathtt{d} = 1$ case for any number of plants N, and for the $\mathtt{d} = 2$ case we know how to do it when the number of plants is "large". Interestingly, the limit that we explore allows us to apply both results. For the heterogeneous space case, Proposition 1 provides our key result. The proposition establishes that we can use a "large N" limit to obtain a simple characterization of the firm's optimization problem when space is heterogeneous. The key insight is that when economic activity is continuous over space, it is *locally* uniform. As a result, in the limiting economy, the solution for homogeneous space applies locally. The proposition states that, in the

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

limit, the optimal policy is to place plants so that local catchment areas are uniform, infinitesimal, intervals in one-dimensional space and hexagons in two-dimensional space. 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.

Section 2.4 showed that, when economic activity is uniform, the solution for one-dimensional space is simpler than with two dimensions. When economic activity is heterogeneous across space, the problem is considerably more complex for either one or two dimensions, as discussed in Section 2.1.1. The *local* logic that we exploit in this paper allows us to make substantive progress in both one- and two-dimensional problems with spatial heterogeneity: Once we know the solution for homogeneous space for the "large N" limit, which we have from Fejes Toth (1953), working with two-dimensional space is no more difficult than working in one dimension. For readability, we describe the proof when space is two-dimensional.

2.5.1 A Sketch of the Proof of Proposition 1

In economy Δ, firm j's profit is given by

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

The strategy is to create upper and lower bounds for firm j's profit. We start by dividing the space $\mathcal{S}$ 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 Δ, k, we construct upper bound and lower bounds on firm j's profit, $\bar{\pi}_j^{k\Delta}$ and $\underline{\pi}_j^{k\Delta}$, such that

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

To construct the upper bound, we begin by considering a best-case scenario for each square by supposing that each square's highest demand, highest effective productivity, and lowest rent apply everywhere in the square. That is, for location s in square $\mathcal{S}_i^k$, we replace D_s, b_s, and R_s with $\bar{D}_s^k \equiv \sup_{\tilde{s} \in \mathcal{S}_i^k} D_{\tilde{s}}$, $\bar{b}_s^k \equiv \sup_{\tilde{s} \in \mathcal{S}_i^k} b_{\tilde{s}}$, $\underline{R}_s^k \equiv \inf_{\tilde{s} \in \mathcal{S}_i^k} R_{\tilde{s}}$. Similarly, to construct the lower bound, we consider a worst-case scenario for each square by supposing that each square's lowest demand, lowest effective productivity, and highest rent apply everywhere in the square.

In Appendix G we explore an example where we solve numerically for the upper and lower bounds $\bar{\pi}_j^{k\Delta}$ and $\underline{\pi}_j^{k\Delta}$. Intuitively, the example shows that the bounds get much tighter for small Δ's, when the chosen number of local plants is large.

We next use the Sum of Moments theorem separately for each square to give an upper bound on the profit the firm could attain from that square in that best-case scenario. As described in Figure 6, if the firm chooses to place N_i plants in square $\mathcal{S}_i^k$, the upper bound corresponds to assigning to each of those plants a catchment area that is a regular hexagon with area $\frac{k \times k}{N_i}$.[17] To construct a lower bound for the

[17] In constructing this upper bound, we compute, separately for each square $\mathcal{S}_i^k$, the profits the firm would earn from the plants

worst-case scenario, we impose an ad hoc restriction on the firm's strategies so that all plants within $\mathcal{S}_i^k$ have catchment areas that are regular hexagons of the same size and that are fully contained in the square $\mathcal{S}_i^k$; since regular hexagons do not form a tesselation 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 fix an arbitrary k and study the limit as $\Delta \to 0$. Define the function $\kappa(n_s) \equiv n_s g(1/n_s)$. We prove 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

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

If we define $\pi_j \equiv \lim_{\Delta\to 0} \pi_j^\Delta$ to be the firm's profit in the limiting economy, these together with $\underline{\pi}_j^{k\Delta} \leq \pi_j^\Delta \leq \bar{\pi}_j^{k\Delta}$ imply that

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

We obtain these results because, for any k, the economic features are uniform within each square $\mathcal{S}_i^k$ in both the best and worst case scenarios, 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

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

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.

2.6 Convergence of the Policy Function

Proposition 1 established that the firm's profit function converges to a well-behaved limit. Here we establish further that the policy function, the choice of the measure of plants by location, converges as well.

Proposition 2 *Suppose that the problem in the limiting economy has a unique solution, n^*. For each Δ, let $O^{\Delta *}$ be a solution to the problem in economy Δ. Then for any ϵ, there is a $\bar{\Delta}$ small enough so that for*

in that square if it had no other plants in all of $\mathcal{S}$. We then add together the profits for all $k \times k$ squares. This double-counting of customers is fine because we are constructing an upper bound for profits, not the value of profit for any feasible policy. In any case, as $\Delta \to 0$, sales outside a plant's $k \times k$ square go to zero.

any $\Delta < \bar{\Delta}$ and any Jordan-measurable set $\mathcal{A} \subseteq \mathcal{S}$,

$$\left| \Delta^{\mathsf{d}} \left| O^{\Delta *} \cap \mathcal{A} \right| - \int_{s \in \mathcal{A}} n_s^* ds \right| < \epsilon$$

The proposition describes the precise sense in which the policy function converges. Consider any Jordan-measurable set of locations, $\mathcal{A}$.[18] As $\Delta \to 0$, the number of plants a firm places in a set of locations naturally rises, as rental costs fall and trade costs rise. Nevertheless, appropriately scaled by Δ^{d}, the number of plants placed in the set converges to a well-behaved limit that corresponds to the solution in the limit economy.[19] Hence, the proposition suggests a natural way to approach data on plant locations. Namely, rather than asking *whether* a firm placed a plant in a particular location, the proposition suggests looking at the number of plants a firm places in a contiguous area (e.g., a 12×12 mile square).

2.6.1 A Sketch of the Proof of Proposition 2

We show uniform convergence of the policy function in two steps. First, we derive properties of a firm's limiting problem. We show that if the limiting problem has a unique solution, n^*, then for any $\varepsilon > 0$ there exists an $\eta > 0$ such that $n \in \bar{\mathcal{N}}$ and $|\Pi(n) - \Pi(n^*)| < \eta$ imply $\int_{s \in \mathcal{S}} |n_s - n_s^*| \, ds < \varepsilon$, where $\bar{\mathcal{N}}$ is a space of functions with a uniform bound and $\Pi(n)$ is the profit the firm would obtain from following strategy n.

In the second step, we study the sequence of economies as $\Delta \to 0$. As in the proof of convergence of the value function in Appendix A.2, we construct a sequence of bounds on the profit function that get tighter as $\Delta \to 0$. We show that for economy Δ, the optimal choice $O^{\Delta *}$ has a corresponding strategy in the limiting economy, $n^{\Delta *}$. As $\Delta \to 0$, the bounds get tighter and two things happen. First, $O^{\Delta *}$ gets close to $n^{\Delta *}$: over any Jordan measurable set $\mathcal{A}$, $\Delta^d \left| O^{\Delta *} \cap \mathcal{A} \right|$ uniformly approaches $\int_{s \in \mathcal{A}} n_s^{\Delta *} ds$. Second, the corresponding strategy $n^{\Delta *}$ delivers a value in the limit economy close to optimum. This, along with the first step, implies that $n^{\Delta *}$ converges to n^*. Namely, we have uniform convergence of the policy function to n^*.

In constructing the strategy $n^{\Delta *}$, we use segments of length k or $k \times k$ squares to find upper and lower bounds, as in the proof of Proposition 1. In particular, for any Δ and k we can construct a strategy in the limiting economy that maximizes profit subject to the restriction that the measure of plants on each segment/square corresponds to the number of plants of $O^{\Delta *}$ in the segment/square multiplied byΔ^{d}. In the

[18] Jordan-measurable sets are, loosely, those that are well-approximated by finite unions of rectangles. These include all bounded convex sets, but not all Lebesgue-measurable sets. A set is Jordan-measurable if and only if its indicator function is Riemann integrable. Why is the theorem restricted to Jordan-measurable sets? The set $O^{\Delta *}$ is finite for any $\Delta > 0$ and thus always has Lebesgue measure zero. It is hard to rule out the possibility that a single set $\mathcal{A}$ with Lebesgue measure zero (e.g., points in the unit square with rational coordinates) contains $O^{\Delta *}$ for all Δ, in which case the Lebesgue integral $\int_{s \in \mathcal{A}} n_s^* ds$ would equal zero.

[19] We do not know if there is a unique optimal policy function for economy Δ (and we have no way of checking). However, the theorem applies to *any* optimal policy functions $O^{\Delta *}$. It is easier to assess uniqueness for the limiting economy. We can divide the problem of the limiting economy into a one-dimensional problem of choosing the total measure of plants, and a sub-problem of choosing a spatial allocation of those plants. It is straightforward to show that conditional on N, there is a unique solution to the sub-problem of placing the plants in space (up to sets of measure zero). We do not provide sufficient conditions to ensure uniqueness of the outer problem, but the fact that it is one-dimensional means that it is easy to verify uniqueness numerically.

proof of Proposition 1, the key step was to take the limit as $\Delta \to 0$ for a given k and then take $k \to 0$. Here, the key trick is to choose a sequence of $k = K(\Delta)$ so that, as we take the limit as $\Delta \to 0$, the sequence $k = K(\Delta)$ also converges to zero (albeit more slowly than does Δ). As such, for each Δ, we construct the strategy $n^{\Delta *}$ in the limit economy that maximizes profits subject to the restriction that the measure of plants on each $K(\Delta)$ segment, or $K(\Delta) \times K(\Delta)$ square, corresponds to the number of plants of $O^{\Delta *}$ in the segment/square multiplied by $\Delta^{\mathtt{d}}$.

2.7 The Local Efficiency of Distribution and its Properties

As we discussed before, we refer to the function $\kappa(n_s) \equiv n_s g(1/n_s)$ as the *local efficiency of distribution* in the neighborhood of s. Recall that $\kappa(n_s)$ represents the fraction of the value of local sales a firm retains after subtracting the cost of optimally transporting the goods to consumers from its n_s plants. The following lemma describes some useful properties of κ.[20]

Lemma 3 $\kappa(n) \equiv ng\left(\frac{1}{n}\right)$ *is strictly increasing and strictly concave, and satisfies the following properties:*

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

If transport costs satisfy $\lim_{\delta\to\infty} \delta^{-\frac{2\mathtt{d}}{\varepsilon-1}} t(\delta) = \infty$ *then*

4. $\kappa''(0) = 0$;
5. $\kappa'(0) < \infty$.

The first property says that with no plants revenues are zero. $\kappa(n)$ is increasing since more plants imply that customers are on average closer to a plant. It is concave since additional plants cannibalize existing plants leading to diminishing gains from reducing transport costs. The second property states that as n grows infinitely large, catchment areas grow small and $\kappa(n)$ approaches an upper bound of 1; in the limit, additional plants provide no significant gains, and 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 provide 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. Put together, cannibalization is irrelevant for the first set of plants but

[20] When space is one dimensional, we can prove a converse of Lemma 3. If κ is twice continuously differentiable, strictly increasing, strictly concave, and satisfies $\kappa(0) = 0$, $\kappa'(0) \in (0, \infty)$, $\kappa''(0) = 0$, $\lim_{n\to\infty} n[1 - \kappa(n)] \in (0, \infty)$, then there is a strictly increasing transport cost $t(\delta)$ that generates κ, namely $t(\delta) = \left[1 + \int_{\frac{1}{2\delta}}^{\infty} \kappa''(x) x dx\right]^{\frac{1}{1-\varepsilon}}$. For example, $\kappa(n) = \frac{\tan^{-1}(\phi n)}{\pi/2}$ for some constant $\phi > 0$ is consistent with the trade cost $t(\delta) = \left\{\frac{1}{\pi/2}\left[\tan^{-1}\left(\frac{1}{2\delta\phi}\right) - \frac{1}{2\delta\phi + (2\delta\phi)^{-1}}\right]\right\}^{\frac{1}{1-\varepsilon}}$.

becomes 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, $n_s = 0$.[21]

2.8 The Assignment of Plants to Locations

Proposition 1 can be used to characterize 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 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

$$x_s z_j^{\varepsilon - 1} \kappa'(n_{js}) \leq R_s + \lambda_j, \qquad \text{with equality if } n_{js} > 0, \tag{4}$$

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

$$\lambda_j = -\frac{d[z(q_j, N_j)^{\varepsilon - 1}]}{dN_j} \int_s x_s \kappa(n_{js}) ds. \tag{5}$$

The productivity of the firm declines with its span of control, as measured by the total number of plants, N_j. Hence, λ_j can be interpreted as the marginal span-of-control cost for the firm. It amounts to an additional shadow fixed cost, on top of the explicit fixed cost R_s, of operating one more plant in location s.

Since κ is strictly concave, (4) implies that n_{js} is increasing in x_s and z_j, and decreasing in R_s and λ_j. $R_s + \lambda_j$ comprise a plant's effective fixed cost. Naturally, higher effective fixed costs induce the firm to operate fewer plants in a location. The firm trades off this effective fixed cost against the gains from increasing the efficiency of distribution: more plants implies that, on average, customers will be closer to the plants. Larger x_s or z_j imply larger gains from a reduction in average distance, inducing the firm to operate more plants in the location.

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 (4), together with this assumption, we can show that firms with higher

[21] In one dimension, $\mathtt{d} = 1$, $\kappa'(0) = 2\int_0^\infty t(\delta)^{1-\varepsilon} d\delta < \infty$. In two dimensions, $\mathtt{d} = 2$, $\kappa'(0) = \int_0^\infty t(\delta)^{1-\varepsilon} 2\pi\delta d\delta < \infty$.

endogenous productivity have higher marginal span-of-control cost of increasing the number of plants, λ_j, even relative to their firm-specific profitability, $z_j^{\varepsilon-1}$.[22]

Lemma 4 *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 $\frac{\lambda_1}{z_1^{\varepsilon-1}} = (\varepsilon-1)\frac{-z_N(q_1,N_1)}{z(q_1,N_1)}\int_s x_s\kappa(n_{1s})ds < (\varepsilon-1)\frac{-z_N(q_2,N_2)}{z(q_2,N_2)}\int_s x_s\kappa\left(n_{2s}\right)ds = \frac{\lambda_2}{z_2^{\varepsilon-1}}$, 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 5 *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 $n_{2s} > 0$; $R_s < R^*\left(z_1, z_2\right)$ implies that $n_{1s} \geq n_{2s}$, with strict inequality if $n_{1s} > 0$; and $R_s = R^*\left(z_1, z_2\right)$ 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 $R_s > R^*\left(z_1, z_2\right)$ and $n_{2s}, n_{1s} > 0$ then $\kappa'\left(n_{2s}\right) = \frac{R_s+\lambda_2}{z_2^{\varepsilon-1}x_s} < \frac{R_s+\lambda_1}{z_1^{\varepsilon-1}x_s} = \kappa'\left(n_{1s}\right)$ 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 $R_s < R^*\left(z_1, z_2\right)$ is identical. The argument for $R = R^*\left(z_1, z_2\right)$ is trivial. ■

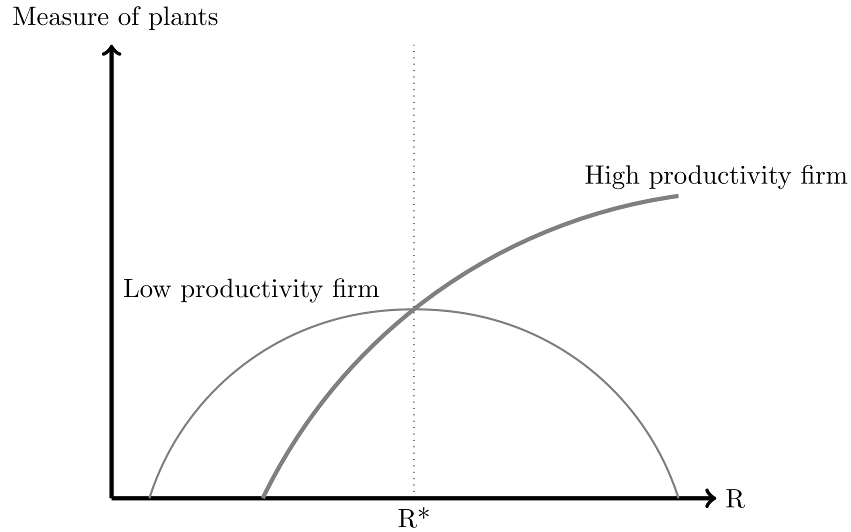
Proposition 5 states that for two firms with different productivities, there is a cutoff level of rent such that the firm with higher productivity places more plants in locations with higher rent and the firm with lower productivity places more plants in locations with lower rent. Thus, even while the two firms have overlapping footprints, there is a clear pattern of sorting. Figure 7 provides a graphical representation of this result.

The type of sorting implied by Proposition 5 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.[23] Here, it is the less productive firms that go to the lower rent locations. Why the difference?

[22]Log concavity ensures that $\frac{-z_N(q,N)}{z(q,N)}$ is non-decreasing.

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

Figure 7: Location of Plants of a High and a Low Productivity Firm



Consider first locations with the cutoff level of rent where two firms with different productivities choose to place the same number of plants. 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. The higher-productivity firm earns more profit per plant in the location, but chooses not to open more plants because of its higher marginal span-of-control costs (as shown in Lemma 4). At locations with higher rent, the higher rent deters the lower-productivity firm from placing many plants, but it has a smaller impact on the higher-productivity firm's effective fixed costs. Hence, the higher-productivity firm places relatively more plants in these high-rent locations. 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. A parallel argument implies that a lower rent induces the lower-productivity firm to place more plants, while the large marginal span-of-control cost of the high-productivity firm limits its presence.[24] Most models of plant location decisions in the literature do not feature span-of-control costs, and so this sorting implication is absent.[25]

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

[24] In the baseline model, a plant requires a fixed amount of space, and production uses only labor. Thus, a plant's fixed cost depends on the local rent and its variable cost depends on the local wage. If both a plant's fixed cost and its production of output used (possibly different) bundles of labor and floor space, firms would still sort across locations. However, rather than sorting according to rent, they would sort according to the cost of the fixed-cost input bundle. That is, larger firms would place plants in locations in which the fixed-cost input bundle was more expensive.

[25] We can also show that the marginal efficiency of distribution of more productive firms is relatively smaller in higher-rent locations. Hence, in higher-rent locations, higher-productivity firms saturate the market relatively more and the cannibalization between plants is larger. We relegate the formal statement of these additional results to Appendix A.4.

costs, 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.

Proposition 6 *If* $\lim_{\delta\to\infty} \frac{\delta^{\mathrm{d}}}{t(\delta)^{\varepsilon-1}} = 0$*, 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* $z_j > \bar{z}_s$*.*

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.[26] We let L_j denote the total number of workers of firm j.

Lemma 7 *Suppose that* $z(q, N) = qe^{-N/\sigma}$ *and local wages are constant across locations at* W*. Consider two firms with* $z_1 < z_2$*, then either* $L_1 < L_2$ *or* $L_1 = L_2 = 0$*.*

3 Industry Equilibrium

We now proceed to embed the problem of the multi-plant firm that we studied in the previous section into an industry equilibrium. We do so for a single 'small' industry in the context of a full spatial equilibrium, of which we do not specify the details.[27] In particular, we are interested in how competition among firms interacts with the sorting of firms across space. After describing the industry equilibrium, we study two comparative statics: a relaxation of the span-of-control cost (perhaps driven by advances in information and communication technologies) and a reduction in transportation costs. We study the implications of these technological developments when the change occurs in 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. The exercises illustrate the relevance of transport costs and span-of-control costs in the limit economy and allow us to speak to the type of changes in sorting documented by Rossi-Hansberg et al. (2021) and Hsieh and Rossi-Hansberg (2022).[28]

[26] The assumption of equal local wages is consistent with the general equilibrium framework in Section 3 of Oberfield et al. (2020)

[27] It is straightforward to embed the industry equilibrium into a full spatial equilibrium framework. This can be done in a number of ways. In one example, spelled out explicitly in Oberfield et al. (2020), each location is characterized by exogenous amenities in addition to productivity, people are freely mobile across locations, and land can be used for housing or commercial real estate. We can also accommodate further agglomeration and congestion forces or impediments to mobility. Of course, other alternative general equilibrium setups could work as well.

[28] We are well-positioned to study this question relative to existing models of plant location which either have no span-of-control cost (Ramondo and Rodríguez-Clare (2013) and Tintelnot (2016)) or limit a firm's location to a single plant (Gaubert (2018) and Ziv (2019)).

There are $\mathcal{L}_s$ workers in location s with Cobb-Douglas preferences across industry aggregates from a unit continuum of industries indexed by $\omega \in [0,1]$. Consistent with Assumption 1, each industry aggregate is a Dixit-Stiglitz bundle of the varieties j produced by all firms in that industry, J_ω, with elasticity of substitution across varieties ε.[29] Thus if $I_{s\omega}$ is the total expenditure on the industry aggregate for industry ω in locations s, the residual demand curve facing firm $j \in J_\omega$ is $I_{s\omega} P_{s\omega}^{\varepsilon-1} p_j^{-\varepsilon}$, where the price index for industry ω is the $P_{s\omega} \equiv \left(\int_{j \in J_\omega} p_{js}^{1-\varepsilon} dj \right)^{\frac{1}{1-\varepsilon}}$. Aggregating across firms, we can characterize a location's industry price index.[30]

Proposition 8 *In the limit when $\Delta \to 0$, the local price index for industry ω is $P_{s\omega} = \frac{\varepsilon}{\varepsilon-1} \frac{W_s}{B_s \mathcal{Z}_{s\omega}}$, where $\mathcal{Z}_{s\omega} \equiv \left(\int_{j \in J_\omega} z_j^{\varepsilon-1} \kappa(n_{js}) dj \right)^{\frac{1}{\varepsilon-1}}$.*

Note that industry productivity in a location, $\mathcal{Z}_{s\omega}$, is the CES aggregate of the firms' effective productivities, z_j, with the weight on a firm's productivity given by its local efficiency of distribution, $\kappa(n_{js})$. Then, the local price index is just the standard CES markup times the 'aggregate' local marginal cost.[31]

3.1 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.

Let transportation costs take the form $t(\delta;\phi) \equiv t(\delta/\sqrt{\phi})$, where ϕ indexes the efficiency of transportation (i.e. a higher ϕ implies lower trade costs for a given distance traveled, δ). This implies that $\kappa(n;\phi) \equiv \kappa(\phi n)$.[32] 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 $\sigma = 1$. Finally, we posit a one-to-one mapping between locations' total expenditure on the industry $I_{s\omega}$ and rent R_s; that these expenditures are distributed according to a truncated Pareto distribution, and that the distribution of firm productivities q_j is also given by a truncated Pareto distribution.[33]

[29] For simplicity we abstract from firm entry and hold the set of firms fixed in our comparative statics exercise.

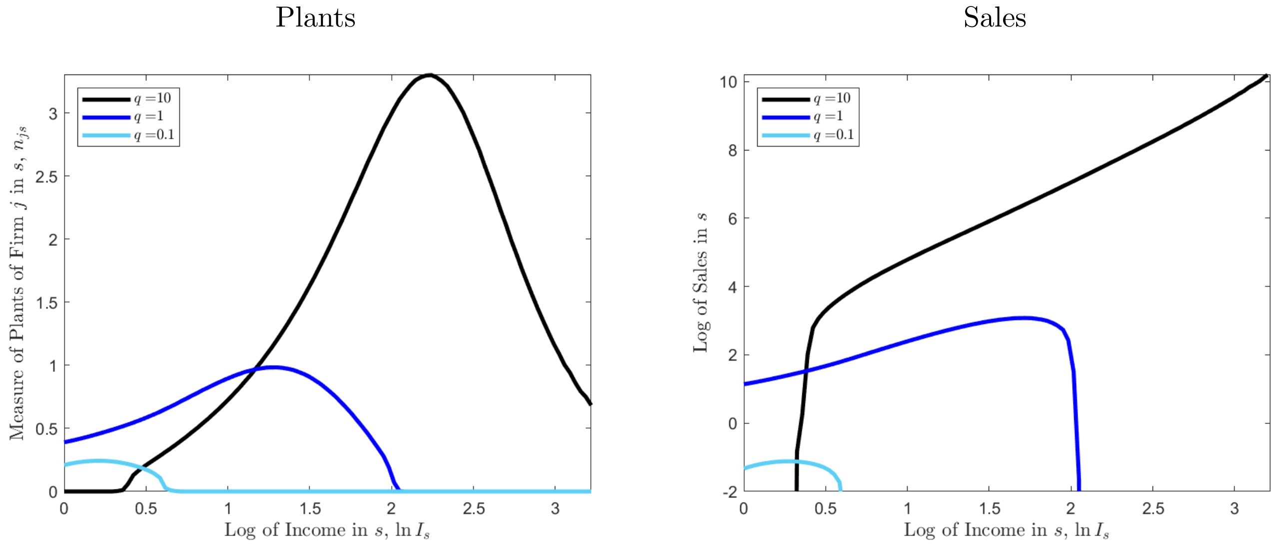
[30] In Appendix A.5 we determine other aggregate properties of the industry equilibrium when $\Delta \to 0$.

[31] We have not been able to show existence or uniqueness of the type of equilibrium we desire. The main holdup is that we have not been able to show that the industry price index in the limiting economy is continuous. While we strongly suspect this is the case—individual firms have incentives to place plants where other firms have not—we have no formal proof. We hope future work can improve on this.

[32] 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.

[33] 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 distribution of fundamentals is such that the rent schedule in a location with income I_s is given by $R(I_s) = e^{\log(I_s)^2}$. 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 8: Sorting in Industry Equilibrium



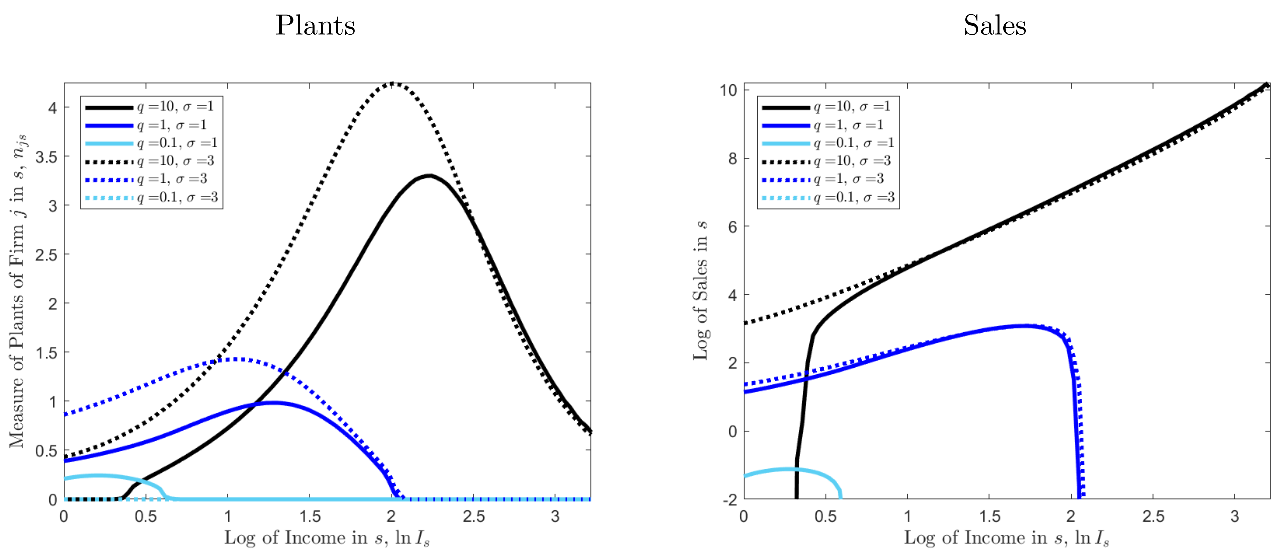
We first describe the baseline industry equilibrium and then proceed to study comparative static exercises with respect to σ and ϕ. Appendix F describes the numerical algorithm that solves the industry equilibrium. Figure 8 presents the distribution of plants, n_{js}, and sales, $(\varepsilon-1)z_j^{\varepsilon-1}x_s\kappa(n_{js})$, 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 5, 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 8). In fact, it operates no plants in the lowest-income 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 8, they compensate with higher sales from each plant which results in higher total sales. Low-income locations, in contrast, are less attractive to large 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-rent locations given their lower span-of-control and the lack of competition from top firms in those locations.

3.1.1 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 8 (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.

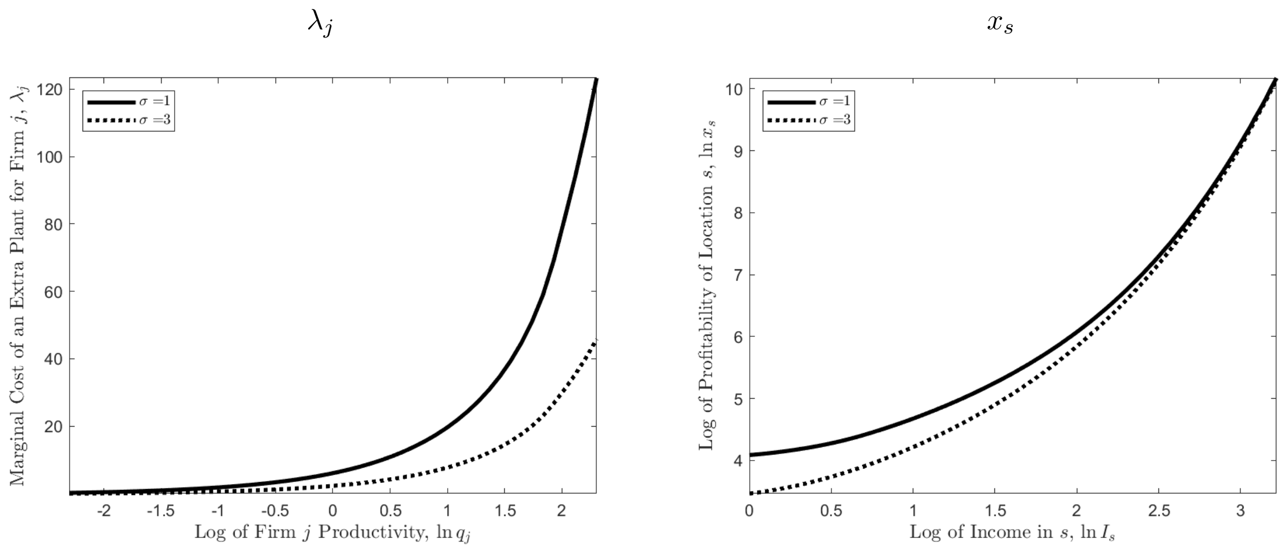
Figure 9: span-of-control and Sorting in an Industry Equilibrium



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 the local price index, particularly in low-income locations, which, in turn, lowers local profitability. 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

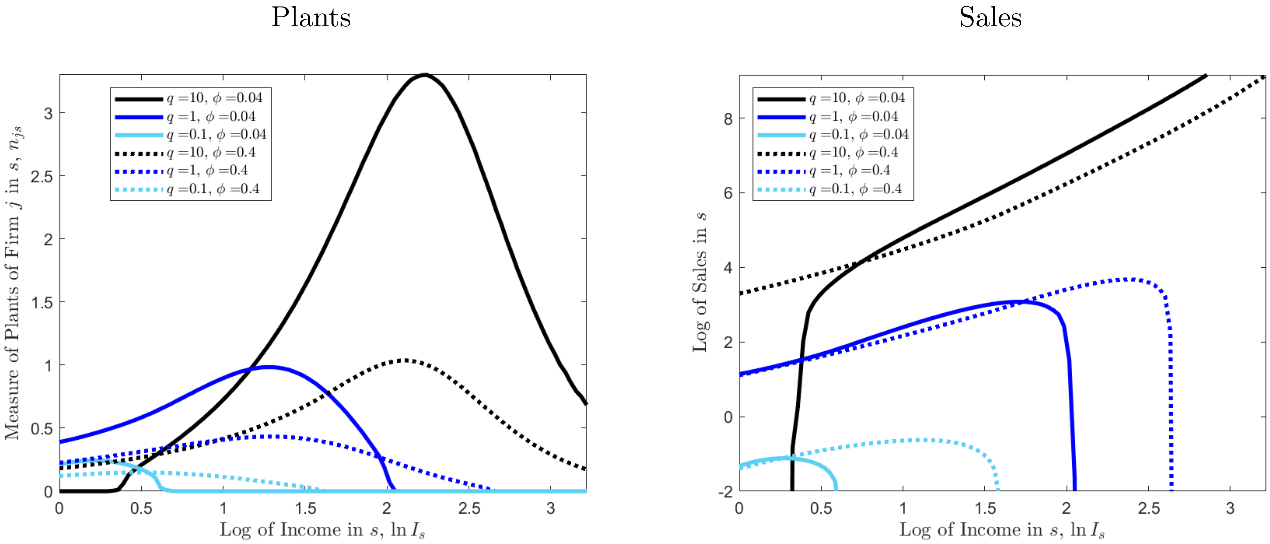


3.1.2 Improvements in an Industry's Transportation Costs

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 imply 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.

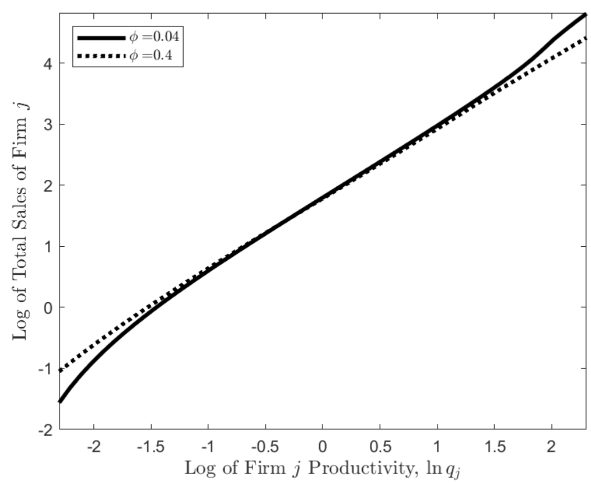
Figure 11 shows the effect of an increase in ϕ from 0.04 to 0.4 on the mass of plants and sales of the representative high, medium, and low productivity firms discussed earlier. The left-hand panel shows that all firms expand to new locations but also have fewer plants in most locations where they were 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

Figure 11: Transportation Efficiency and Sorting in an Industry Equilibrium



competition implies that all three firms see their sales fall in many of the markets where they were operating.

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



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

4 Empirical Evidence

Our theory provides a number of concrete implications about the location of plants in space for industries that can be approximated by our limit economy. In this section, we contrast its implications on sorting and the role of span-of-control with U.S. evidence. 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 remainder of the paper.[34] For each establishment, we know its geographic coordinates, its industry classification, and its parent company.[35] 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 lengths of M miles. We present results for values of M ranging from 3 to 48 miles.

In order to contrast the theory's predictions with the data, we first need to map firm productivity and location characteristics to observable measures in the data. In Lemma 7, 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.[36] We can easily measure each location's population density in the data (since all locations are squares with the same area), and then use this metric to rank locations.

We have characterized the model's predictions for the limit economy. While the underlying forces we underscore, such as transport costs, span-of-control, and cannibalization are likely relevant for any multiplant firm, the predictions of our theory are only guaranteed to hold in the limit economy. This limit is likely a better approximation for firms that set many plants across locations. Thus, when assessing the model's predictions, we do so both using all industries and also restricting attention to industries in which plant catchment areas tend to be small, like services.[37]

[34]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 business. 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.

[35]A more detailed description of NETS can be found in Rossi-Hansberg et al. (2021). We 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, D, and E, we show that our empirical findings regarding sorting and span-of-control are robust to using only the non-imputed data.

[36]In our data, employment is better measured than revenue. Note also that, in our data, franchises are listed as separate firms. We hope future research can explore how span-of-control considerations are affected or relaxed by contracting arrangements such as franchise agreements.

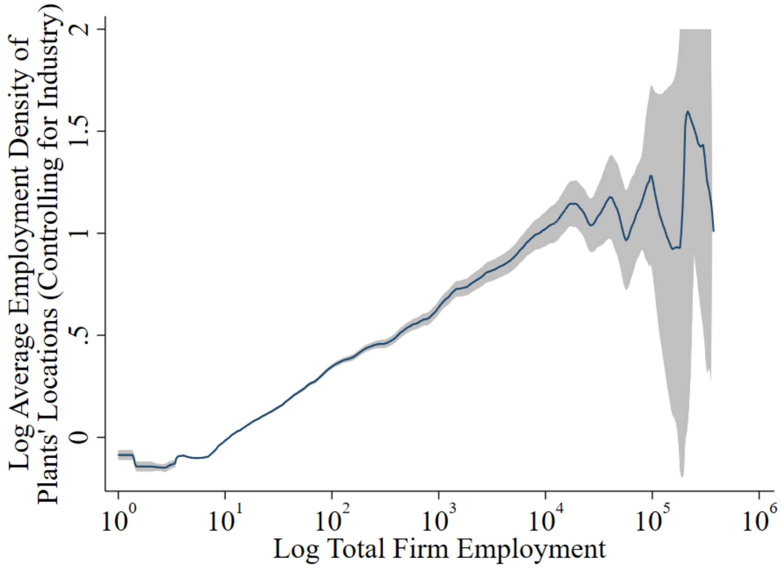
[37]In Appendix H we show, for a numerical example, that the limit predictions derived from our model hold well for simulated plant locations when Δ is small, but may fail when Δ is large and firms set up very few plants across space.

4.1 Sorting in the Data

A central and distinctive prediction of our model is that more productive firms sort towards 'better' locations. Our main results related to sorting are presented in Proposition 5. This proposition establishes that more productive firms set up relatively more plants in locations with higher land rents. 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 data to estimate densities and Zillow data to compute rents. 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 density of the locations of a firm j, $\bar{L}_j$, is the average of the location density, $\mathcal{L}_s$, across all of the firm's plants. Once we compute $\bar{L}_j$, we subtract industry fixed effects. We use the residuals as our measurement of a firm's average density of its locations. The results below are robust to constructing $\bar{L}_j$ using a weighted average, where the weights are based on total employment in the plant's location.

Figure 13: Sorting: Firm Size and Local Density



Notes: For each firm, we calculate the log of average employment density across all of the firm's plants (we use $M = 12$). Then, we subtract industry fixed effects and collect the residuals. Finally, we fit a kernel-weighted local regression of the residuals on the log of total firm employment. The regression shown in the figure uses a zero-degree polynomial (local-mean smoothing) and the bandwidth that minimizes the conditional weighted mean integrated squared error. The shaded area indicates the 95% percent confidence interval.

Figure 13 shows that the relationship between $\ln \bar{L}_j$ and $\ln L_j$ is roughly linear when we restrict attention to firms with national size in an industry greater than 10 employees. Hence, in Table I we estimate a linear relationship and show that, indeed, the relationship between a firm's log average location density and its log national employment size is positive and significant, after controlling for industry fixed effects. The table also presents a selected set of robustness checks. The implication of the theory holds robustly in the data: bigger firms sort towards dense locations. The second column of Table I confirms the same finding when we look at a larger spatial resolution $M = 48$, although the coefficient is smaller probably due to spatial averaging across markets.

The implications we derived from the limit problem where transport costs are large and span-of-control and fixed costs are small should describe particularly well the behavior of firms that choose to set up many plants or that operate in industries in which trade costs are high. Column 3 Table I shows that the sorting pattern is indeed present and strongly significant when we limit the sample to firms with 100 or more plants, even though it reduces the sample size tremendously. Column 4 restricts the sample to industries that have at least one firm with at least 100 plants and finds an even steeper positive relationship which is again highly significant. Finally, 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 a positive relationship between average firm density and total firm employment.[38] We address this concern by examining sorting patterns among firms with the same headquarter locations. As shown in Column 5 in Table I, the positive relationship between total firm employment and the average employment density of the firm's plant locations is robust to the inclusion of fixed effects for the firm's headquarter locations for firms with more than 100 plants.

Table V in Appendix C shows many more variations of these results using different thresholds and selection criteria. All of them show similar findings. In addition, we implement a leave-out strategy to address the potential concern that the firm's presence could be driving local density. The resulting sorting is virtually identical. We also present results when we limit the sample to plants with non-imputed employment data, as well as using alternative weights.

Table II presents the estimated elasticity of firm average location density to firm national employment by major industrial sector. As discussed above, the limit problem we study is likely a better approximation of the problem of firms that sell goods and services at short distances. Broad industry classifications do not provide an ideal grouping of industries according to their tradability. Nevertheless, it is probably the case that firms in industries within, say, retail trade sell services that are less tradable than firms in industries within manufacturing. Table II shows that the elasticity of firm average location density with respect to firm national employment is in fact smaller in manufacturing than in other sectors. We find the highest elasticities in FIRE and services.

4.2 The Largest Firm in Town

We can also explore the implications of Proposition 5 on sorting by looking at how the size of the firm with the largest number of plants in each location, $L_{j^*(s)}$, changes with population density, $\mathcal{L}_s$, where $j^*(s)$ is the identity of the firm that places the most plants in s. Sorting implies that, in locations with low population densities, low productivity and smaller firms should place more plants than large firms. Table III shows that,

[38] See Walsh (2019) for a recent study of firm entry across locations.

Table I: Sorting: Firm Size and Local Density

	Baseline $M = 12$	*Baseline* $M = 48$	*Firms with* ≥ 100 *plants*	*Industries in which largest firm has* ≥ 100 *plants*	*HQ fixed effects, Firms with* ≥ 100 *plants*
	(1)	(2)	(3)	(4)	(5)
	$\ln \bar{L}_j$	$\ln \bar{L}_j$	$\ln \bar{L}_j$	$\ln \bar{L}_j$	$\ln \bar{L}_j$
$\ln L_j$	0.165***	0.0952***	0.146***	0.172***	0.0791**
	(0.000975)	(0.000848)	(0.0249)	(0.00164)	(0.0350)
Observations	3,670,994	3,673,053	876	1,387,742	652
R-squared	0.139	0.099	0.384	0.080	0.664
SIC8 FE	Yes	Yes	Yes	Yes	Yes
HQ Location FE	No	No	No	No	Yes
M	12	48	12	12	12

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

Notes: The table presents the results of regressing the log of the average employment density across all of the firm's plants on the log employment of the firm at the national level and industry fixed effects. The first column presents the baseline results with $M = 12$. The second column shows the baseline the results with $M = 48$. The third column restricts the analysis to firms with at least 100 plants. The fourth column restricts the analysis to industries where there is a firm with at least 100 plants. The fifth column adds headquarters' location fixed effect for each firm to the case where we restrict to firms with at least 100 plants.

Table II: Sorting by Major Industry

By major industry	*All*	*Manufacturing*	*Services*	*Retail Trade*	*FIRE*
	(1)	(2)	(3)	(4)	(5)
	$\ln \bar{L}_j$	$\ln \bar{L}_j$	$\ln \bar{L}_j$	$\ln \bar{L}_j$	$\ln \bar{L}_j$
$\ln L_j$	0.165***	0.0523***	0.160***	0.150***	0.234***
	(0.000975)	(0.00272)	(0.00153)	(0.00247)	(0.00320)
Observations	3,670,994	274,478	1,479,391	856,860	244,048
R-squared	0.139	0.192	0.097	0.068	0.122
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12

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

Notes: The table presents the results of regressing 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, on the log employment of the firm at the national level and industry fixed effects. The first presents the baseline case for all industries, and the rest of the columns present results by major sector.

in fact, the national size of the firm with the most plants in a location increases with population density, controlling for industry fixed effects. The first column presents our baseline case, the second one the case with 48-mile square resolution, the third when we restrict the sample to firms with at least 100 plants, and the fourth when we restrict the sample to industries that have at least a firm with 100 or more plants. In all cases, the slope is positive a highly significant.

Table VI in Appendix D presents a large set of robustness checks, including different thresholds for sample selection, results for major industry groupings, as well as additional spatial resolutions and an exercise with only non-imputed data. In the analysis presented in Table III, there are locations in which multiple firms tie for the highest number of plants. In those cases, we use the average national firm size among these firms. In Table VI in Appendix D we also show that our finding is robust to dropping cases with ties or to using the national size of the largest firm among those tied. Finally, one may worry that the largest firm in a location could be large enough to mechanically and significantly affect the local employment density. In Table VI we show that the results are only marginally affected when excluding the firm's own contribution when calculating local employment density.

Table III: National size of the largest firm in town

	Baseline $M = 12$	*Baseline* $M = 48$	*Firms with* ≥ 100 *plants*	*Industries in which largest firm has* ≥ 100 *plants*
	(1)	(2)	(3)	(4)
	$\ln L_{j^*(s)}$	$\ln L_{j^*(s)}$	$\ln L_{j^*(s)}$	$\ln L_{j^*(s)}$
$\ln \mathcal{L}_s$	0.395***	0.594***	0.131***	0.516***
	(0.00266)	(0.00461)	(0.00811)	(0.00366)
Observations	1,984,474	1,006,305	211,517	616,248
R-squared	0.616	0.644	0.554	0.600
SIC8 FE	Yes	Yes	Yes	Yes
M	12	48	12	12

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

Notes: The table presents the results of finding the log employment of the firm with the most plants in an industry and location, and regressing its log total employment on the log density of the location and 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. Column 1 presents the baseline results when $M = 12$ and column 2 for $M = 48$. Column 3 restricts the analysis to firms with at least 100 plants, and column 4 restricts the analysis to industries where there is a firm with at least 100 plants. For population density, we use the data from the 2010 decennial census taken from Manson et al. (2021).

4.3 The Role of Span-of-control Costs

In this section, we present evidence on a particular mechanism driving firm sorting across locations in our model. Lemma 4 shows that higher productivity firms have a higher cost of increasing their span-of-control by an additional plant, λ_j. As is evident from equation 4, in choosing the number of plants in a given location, firms trade-off these firm-specific fixed costs against profits per plant, which are increasing in a firm's productivity. This trade-off implies that two firms present in the same location, but with different productivity, might decide to have the same number of plants. However, the firm with higher productivity will always have larger plants. Hence, a testable prediction of this mechanism is that, among firms with the same number of plants in a given location, the plants operated by the more productive, and therefore nationally larger, firm should be larger.

Formally, we can write the average plant size of firm j in location s as

$$\bar{l}_{js} = (\varepsilon - 1) z_j^{\varepsilon - 1} \frac{x_s}{W_s} \frac{\kappa(n_{js})}{n_{js}}. \tag{6}$$

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 $n_{js} = n_{\tilde{j}s}$. Note that, in most models of multi-plant production, a firm's effective productivity in a given location (for example, its productivity adjusted by the location's distance to the firm's headquarters) determines both the number of plants and the size of those plants. In contrast to our prediction, this implies that there should be no systematic relationship between firm productivity and plant size after controlling for the number of plants in a location.

Table IV: Span of control

	Baseline $M = 12$	*Baseline* $M = 48$	*Firms with* ≥ 100 *plants*	*Industries in which largest firm has* ≥ 100 *plants*
	(1)	(2)	(3)	(4)
	$\ln \bar{l}_{js}$	$\ln \bar{l}_{js}$	$\ln \bar{l}_{js}$	$\ln \bar{l}_{js}$
$\ln L_{j,-js}$	0.114***	0.131***	0.275***	0.104***
	(0.000897)	(0.000962)	(0.00296)	(0.000911)
$\ln n_{js}$	0.137***	0.172***	-0.168***	0.0720***
	(0.00897)	(0.00682)	(0.0111)	(0.00883)
$(\ln n_{js})^2$	-0.0813***	-0.0811***	-0.00158	-0.0564***
	(0.00447)	(0.00251)	(0.00528)	(0.00431)
Observations	409,364	386,094	126,999	336,424
R-squared	0.573	0.511	0.746	0.588
SIC8-location FE	Yes	Yes	Yes	Yes
M	12	48	12	12

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

Notes: The table presents the results of regressing the log of the average plant employment of a firm within a location on the log national employment of the firm (excluding the own firm contribution of employment in a location from that firm's total employment), industry fixed effects and controls for the number of plants that the firm has in the location. Column 1 presents the baseline results when $M = 12$ and column 2 for $M = 48$. Column 3 restricts the analysis to firms with at least 100 plants, and column 4 restricts the analysis to industries where there is a firm with at least 100 plants.

Table IV presents our estimates of the relationship between log average plant employment, $\bar{l}_{js}$, and the log of total firm employment in alternative locations, $L_{j,-js}$, after controlling for the number of plants and industry-location fixed effects. To control for the term $\kappa(n_{js})/n_{js}$ we use a second-order polynomial in $\ln n_{js}$. We exclude the location's employment from a firm's total employment in order to avoid a mechanical relationship between national firm size and local average plant size. In Table VIII in Appendix E we show that the results are similar if we use a higher order polynomial to approximate $\kappa(n_{js})/n_{js}$, or if we calculate a firm's national size including the location's employment. As before, in Table IV we present estimates for different spatial resolutions, as well as for samples where we restrict attention to firms with more than 100 plants, or to industries that have such firms. In all cases the relationship between average plant size and national firm size is positive and significant, after controlling for the number of establishments.[39]

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 simpler calculus of variations problem. The solution can be easily characterized and the problem can be readily incorporated into a general equilibrium spatial setup with labor mobility.

The solution to the firm's problem has a number of unique predictions. First, and most important, is that span-of-control considerations imply that firms sort in space. Specifically, more productive firms operate relatively more plants in locations with higher rents. Less productive firms, in turn, operate more plants in 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 2014 in the U.S., both when we look at all industries and when we restrict the sample to industries with small catchment areas that might be better approximated by the limit economy.

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. We also 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.

[39] 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., the ones that rely on the distributional assumptions discussed in Lind and Ramondo (2018).

References

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., Eckert, F., and Shi, R. (2017). Combinatorial discrete choice. Technical report.

Arkolakis, C., Ramondo, N., Rodríguez-Clare, A., and Yeaple, S. (2018). Innovation and production in the global economy. *American Economic Review*, 108(8):2128–73.

Balinski, M. L. (1965). Integer programming: methods, uses, computations. *Management science*, 12(3):253–313.

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

Bilal, A. and Rossi-Hansberg, E. (2021). Location as an asset. *Forthcoming in Econometrica.*

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

Byrka, J. and Aardal, K. (2010). An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem. *SIAM Journal on Computing*, 39(6):2212–2231.

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.*

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.

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.

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.

Fowler, R. J., Paterson, M. S., and Tanimoto, S. L. (1981). Optimal packing and covering in the plane are np-complete. *Information processing letters*, 12(3):133–137.

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

Gersho, A. (1979). Asymptotically optimal block quantization. *IEEE Transactions on information theory*, 25(4):373–380.

Guha, S. and Khuller, S. (1999). Greedy strikes back: Improved facility location algorithms. *Journal of algorithms*, 31(1):228–248.

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. (2022). The industrial revolution in services. *Journal of Political Economy Macroeconomics*, forthcoming.

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.

Li, S. (2013). A 1.488 approximation algorithm for the uncapacitated facility location problem. *Information and Computation*, 222:45–58.

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

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., and Ruggles, S. (2021). IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN: IPUMS.

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

Nocke, V. (2006). A gap for me: Entrepreneurs and entry. *Journal of the European Economic Association*, 4(5):929–956.

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

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.

Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2021). Diverging trends in national and local concentration. *NBER Macroeconomics Annual*, 35:115–150.

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

Stollsteimer, J. F. (1961). *The effect of technical change and output expansion on the optimum number, size, and location of pear marketing facilities in a California pear producing region.* University of California, Berkeley.

Sviridenko, M. (2002). An improved approximation algorithm for the metric uncapacitated facility location problem. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 240–257. Springer.

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

Walsh, C. (2019). Firm creation and local growth. Technical report, http://dx.doi.org/10.2139/ssrn.3496782.

Weber, A. (1909). Über den standort der industrien, tübingen theory of the location of industries.

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

Online Appendix

A Proofs

A.1 Proof of Lemma 3: Properties of the Efficiency of Distribution

We provide a proof of Lemma 3 in this section before proving Propositions 1 and 2 in the next two sections, even though the latter come earlier in the main text. We do this because Lemma 3 does not depend on those propositions, but the fact that $\kappa(\cdot)$ is increasing and concave is useful in the proofs of those propositions.

We first prove Lemma 3 for the case in which space is one dimensional, followed by the case in which space is two dimensional.

A.1.1 One Dimension

First, define $\tilde{t}(\delta) \equiv t(\delta)^{1-\varepsilon}$. The integral of the function $\tilde{t}(|s|)$ over a line segment of length x can be expressed as

$$g(x) = \int_{-x/2}^{x/2} \tilde{t}(|\delta|)\, d\delta = 2\int_0^{x/2} \tilde{t}(\delta) d\delta.$$

It will sometimes be convenient to change variables and express this as

$$g(x) = x\int_0^1 \tilde{t}\left(\frac{x}{2}u\right) du.$$

We thus have two expressions for the efficiency of distribution:

$$\kappa(n) = ng\left(\frac{1}{n}\right) = 2n\int_0^{\frac{1}{2n}} \tilde{t}(\delta) d\delta \tag{7}$$

$$\kappa(n) = ng\left(\frac{1}{n}\right) = \int_0^1 \tilde{t}\left(\frac{u}{2n}\right) du \tag{8}$$

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

$$\kappa'(n) = 2\int_0^{\frac{1}{2n}} \tilde{t}(\delta) d\delta - \frac{1}{n}\tilde{t}\left(\frac{1}{2n}\right) \tag{9}$$

$$\kappa'(n) = \int_0^1 \left[-\frac{1}{2n^2}\tilde{t}'\left(\frac{u}{2n}\right)\right] du \tag{10}$$

An expression for the second derivative comes from differentiating (9)

$$\kappa''(n) = \frac{1}{2n^3}\tilde{t}'\left(\frac{1}{2n}\right) \tag{11}$$

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 1 as $n \to \infty$, i.e., $\lim_{n\to\infty} n\,[1 - \kappa(n)] = -\frac{1}{4}\tilde{t}'(0) > 0$.*

Proof. $\tilde{t}$ is strictly decreasing because t is strictly increasing and $\varepsilon > 1$. As a result, (10) implies that κ is strictly increasing and (11) implies that κ is strictly concave. $\kappa(0) = 0$ follows from (8) 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 (8) and $\tilde{t}(0) = 1$.

Beginning with (8), taking the limit as $n \to \infty$, using $x = 1/n$, and applying L'Hopital's rule gives

$$\begin{aligned}
\lim_{n\to\infty} n\,(1 - \kappa(n)) &= \lim_{n\to\infty} n\left(1 - \int_0^1 \tilde{t}\left(\frac{u}{2n}\right) du\right) \\
&= \lim_{x\to 0} \frac{1 - \int_0^1 \tilde{t}\left(\frac{ux}{2}\right) du}{x} \\
&= \lim_{x\to 0} \frac{0 - \int_0^1 \frac{u}{2}\tilde{t}'\left(\frac{ux}{2}\right) du}{1} \\
&= -\frac{1}{4}\tilde{t}'(0)
\end{aligned}$$

■

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

Proof. Taking the limit of the (9) gives

$$\begin{aligned}
\lim_{n\to 0} \kappa'(n) &= \lim_{n\to 0} 2\int_0^{\frac{1}{2n}} \tilde{t}(\delta)d\delta - \frac{1}{n}\tilde{t}\left(\frac{1}{2n}\right) \\
&= 2\int_0^\infty \tilde{t}(\delta)d\delta - 2\lim_{\delta\to\infty} \delta\tilde{t}\,(\delta)
\end{aligned}$$

■

Claim A.3 *If $\lim_{\delta\to\infty} \tilde{t}(\delta)\delta^2 = 0$, then $\kappa''(0) = 0$*

Proof. The second derivative of κ at zero is defined as $\kappa''(0) = \lim_{n\to 0} \frac{\kappa'(n) - \kappa'(0)}{n}$. $\lim_{\delta\to\infty} \tilde{t}(\delta)\delta^2 = 0$

implies $\lim_{\delta\to\infty}\tilde{t}(\delta)\delta = 0$, so Claim A.2 gives $\kappa'(0) = 2\int_0^\infty \tilde{t}(\delta)d\delta$. Using this along (9) gives

$$\begin{aligned}
\kappa''(0) &= \lim_{n\to 0}\frac{\kappa'(n)-\kappa'(0)}{n} \\
&= \lim_{n\to 0}\frac{2\int_0^{\frac{1}{2n}}\tilde{t}(\delta)d\delta + \frac{1}{n}\tilde{t}\left(\frac{1}{2n}\right) - 2\int_0^\infty \tilde{t}(\delta)d\delta}{n} \\
&= \lim_{n\to 0}\frac{\frac{1}{n}\tilde{t}\left(\frac{1}{2n}\right)}{n} + \lim_{n\to 0}\frac{2\int_0^{\frac{1}{2n}}\tilde{t}(\delta)\,d\delta - 2\int_0^\infty \tilde{t}(\delta)\,d\delta}{n}
\end{aligned}$$

Using L'Hopital's rule for the second term and then changing variables to $\delta = \frac{1}{2n}$ gives

$$\begin{aligned}
\kappa''(0) &= \lim_{n\to 0}\frac{1}{n^2}\tilde{t}\left(\frac{1}{2n}\right) + \lim_{n\to 0}\left[-2\frac{1}{2n^2}\tilde{t}\left(\frac{1}{2n}\right)\right] \\
&= 4\lim_{\delta\to\infty}\delta^2\tilde{t}(\delta) + \lim_{\delta\to\infty} -4\delta^2\tilde{t}(\delta) \\
&= 0
\end{aligned}$$

■

A.1.2 Two Dimensions

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 $\tilde{t}(\|s\|)$ 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 $\varpi(1) = 0$.[40] We first rewrite g in a form that is easier

[40]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 fully 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 $\varpi(\delta) = \begin{cases} 1 & 0 \le \delta \le \sqrt{3}/2 \\ 1 - \frac{6}{\pi}\cos^{-1}\left(\frac{\sqrt{3}/2}{\delta}\right) & \sqrt{3}/2 \le \delta \le 1 \end{cases}$.

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

$$\kappa(n) = ng\left(\frac{1}{n}\right) = n\int_0^{\psi n^{-1/2}} \varpi\left(\frac{\delta}{\psi n^{-1/2}}\right)\tilde{t}(\delta)2\pi\delta d\delta \tag{12}$$

$$\kappa(n) = ng\left(\frac{1}{n}\right) = \psi^2\int_0^1 \varpi(u)\tilde{t}\left(\psi n^{-1/2}u\right)2\pi u du \tag{13}$$

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

$$\kappa'(n) = \psi^2\int_0^1 \varpi(u)\tilde{t}'\left(\psi n^{-1/2}u\right)\left(-\psi\frac{1}{2}n^{-3/2}u\right)2\pi u du \tag{14}$$

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

$$\kappa''(n) = \psi^3 n^{-5/2}\frac{\pi}{2}\int_\alpha^1 \varpi'(u)u^3\left[-\tilde{t}'\left(\psi n^{-1/2}u\right)\right]du \tag{15}$$

Claim A.4 $\kappa(n) \equiv ng\left(\frac{1}{n}\right)$ *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.,* $\lim_{n\to\infty}\sqrt{n}\left[1-\kappa(N)\right] = -\tilde{t}'(0)\frac{\sqrt{2}}{3^{3/4}}\left(\frac{1}{3}+\frac{\ln 3}{4}\right) > 0$.

Proof. (14) implies that κ' is strictly positive because $t' > 0$ and $\varepsilon > 1$ imply that $\tilde{t}' < 0$. (15) implies that κ'' is strictly negative because ϖ' is strictly negative on $(\alpha, 1)$. $\kappa(0) = 0$ follows from (13) 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 (13) and the facts that $\tilde{t}(0) = 1$, and $\psi^2 \int_0^1 \varpi(u) 2\pi u du = 1$.

Beginning with (13), 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

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

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 (12) 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

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

Claim A.5 *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 (16) gives

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

The second term of (16) 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 $\lim_{n\to 0}\tilde{t}\left(\psi n^{-1/2}u\right)\left(\psi n^{-1/2}u\right)^2=0$, the limit of the second terms is zero. ■

Claim A.6 *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.5 gives $\kappa'(0)=\int_0^{\infty}\tilde{t}(\delta)2\pi\delta d\delta$. Using this along (16) gives

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

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)

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

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

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

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

A.2 Proof of Proposition 1: Convergence of Profit Function

Define the function $G(x)$ to be the integral of $T\left(\|s\|\right)^{1-\varepsilon}$ over a line segment of length x if $\mathtt{d}=1$ or over a regular hexagon of area x if $\mathtt{d}=2$ centered at the origin. That is,

$$
\begin{aligned}
G(x) &\equiv 2\int_0^{x/2}T(\delta)^{1-\varepsilon}d\delta,\quad \mathtt{d}=1\\
G(x) &\equiv \int_0^{\psi x^{1/2}}\varpi\left(\frac{\delta}{\psi x^{1/2}}\right)T(\delta)^{1-\varepsilon}2\pi\delta d\delta,\quad \mathtt{d}=2
\end{aligned}
$$

where $\varpi(r)$ is the fraction of circle with radius r that intersects with a hexagon with side length 1 as in Appendix A.1.2. Notice that $G(x)=\Delta^{\mathtt{d}}g\left(\frac{x}{\Delta^{\mathtt{d}}}\right)$.[41] We begin by restating a well-known result from discrete geometry.

Theorem A.7 *(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.

[41] This follows from the definition of t along with the change of variables $\tilde{\delta}=\frac{\delta}{\Delta}$. In one dimension, this gives:

$$G(x)=2\int_0^{x/2}t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}d\delta=\Delta 2\int_0^{\frac{1}{\Delta}\frac{x}{2}}t\left(\tilde{\delta}\right)^{1-\varepsilon}d\tilde{\delta}=\Delta g\left(\frac{x}{\Delta}\right)$$

In two dimensions, this gives:

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

The analogous statement in one dimension is straightforward.

Theorem A.8 *Let $f : [0, \infty) \to \mathbb{R}$ be a nondecreasing function and let L be a line segment in $\mathbb{E}$. Then for any set of n points P in $\mathbb{E}$,*

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

where L_n is a line segment in $\mathbb{E}$ with length $|L|/n$ and center at the origin.

Proof. For the n points $\{p_k\}_{k=1}^n$, let ℓ_k be the set of points in the line segment L for which p_k is the closest. Each ℓ_k is a line segment. Let $\bar{x}_k$ and $\underline{x}_k$ denote the upper and lower endpoints of the line segment, and let $x_k^* = \frac{\bar{x}_k + \underline{x}_k}{2}$ denote its center.

Consider the line segment ℓ_k:

$$\begin{aligned}\int_{\underline{x}_k}^{\bar{x}_k} f\left(|x-p_k|\right) &= \int_0^{\bar{x}_k - p_k} f(u)du + \int_0^{p_k - \underline{x}_k} f(u)du \\ &= \int_0^{|\ell_k|/2} f(u)du + \int_{|\ell_k|/2}^{\bar{x}_k - p_k} f(u)du + \int_0^{|\ell_k|/2} f(u)du + \int_{|\ell_k|/2}^{p_k - \underline{x}_k} f(u)du \\ &= G\left(|\ell_k|\right) + \int_{|\ell_k|/2}^{\bar{x}_k - p_k} f(u)du + \int_{|\ell_k|/2}^{p_k - \underline{x}_k} f(u)du \qquad (17)\end{aligned}$$

If $x_k^* \geq p_k$, which implies $\bar{x}_k - p_k \geq |\ell_k|/2 \geq x_k^* - p_k$, the fact that f is increasing implies

$$\int_{|\ell_k|/2}^{\bar{x}_k - p_k} f(u)du \geq \int_{|\ell_k|/2}^{\bar{x}_k - p_k} f\left(u - (x_k^* - p_k)\right) du = \int_{p_k - \underline{x}_k}^{|\ell_k|/2} f(v)dv$$

If, on the other hand, $x_k^* \leq p_k$, which implies $p_k - \underline{x}_k \geq |\ell_k|/2 \geq p_k - x_k^*$, we have

$$\int_{|\ell|/2}^{p_k - \underline{x}_k} f(u)du \geq \int_{|\ell_k|/2}^{p_k - \underline{x}_k} f\left(u - (p_k - x_k^*)\right) du = \int_{\bar{x}_k - p_k}^{|\ell_k|/2} f(v)dv$$

In either case, the sum of the final two terms of (17) are non-negative, giving

$$\int_{\underline{x}_k}^{\bar{x}_k} f\left(|x-p_k|\right) \geq G\left(|\ell_k|\right)$$

Note further that $G(x) = 2\int_0^{x/2} f(u)du$ is convex because f non-decreasing implies that $G'(x) = f\left(\frac{x}{2}\right)$ is non-decreasing. Therefore, Jensen's inequality implies

$$\sum_{k=1}^n G\left(|\ell_k|\right) = n\left(\frac{1}{n}\sum_{k=1}^n G\left(|\ell_k|\right)\right) \geq nG\left(\frac{1}{n}\sum_{k=1}^n |\ell_k|\right) = nG\left(\frac{|L_n|}{n}\right)$$

■

We next apply these theorems to our context.

Lemma A.9 *For any k and any finite set of points $O_i \subset \mathcal{S}_i^k$,*

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

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 and Theorem A.8 therefore imply that

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

■

The next result will be useful in deriving a lower bound for the firm's profit, by studying the profit delivered by a feasible but sub-optimal policy.

Lemma A.10 *For any $k > 0, N \in \mathbb{N}_0$*

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

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

Proof. We first consider the one-dimensional case. It is feasible to place the N points so that the line segment $\mathcal{S}_i^k$ with length k is divided into N segments each of length k/N with an element of O_i at the center of each line segment. Such a choice of O_i would deliver the value $NG\left(\frac{k}{N}\right)$. Since this value is weakly lower than the optimum and G is increasing, we have

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

We next turn two the case of two dimensions. 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 $\left(1 + \frac{1}{\psi\sqrt{N}}\right)^{-2} \frac{k^2}{N}$ inside the square $\mathcal{S}_i^k$ as this would constitute a particular O_i choice. Since the side length of a hexagon with area x has side length of $\psi\sqrt{x}$, 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.11 *For any k, Δ, $\bar{\pi}_j^{k\Delta}\geq\pi_j^{\Delta}$ where*

$$\bar{\pi}_j^{k\Delta}\equiv\sup_{\{N_i\geq 0\}}\sum_{i\in I^k}-N_i\underline{R}_i^k\xi+Z\left(q_j,\sum_{i'\in I^k}N_{i'}\right)^{\varepsilon-1}\sum_{i\in I^k}\bar{D}_i^k\bar{b}_i^kN_iG\left(\frac{k^{\mathtt{d}}}{N_i}\right)+Z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k T(\delta)^{1-\varepsilon}d\delta$$

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 Lemma A.9. 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,

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

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 $\sum_i\sum_{i'\neq i}\bar{D}_i^k\bar{b}_{i'}^k\int_{s\in\mathcal{S}_i^k}\max_{o\in O_{i'}^k} T\left(\delta_{so}\right)^{1-\varepsilon} ds \leq \bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k T(\delta)^{1-\varepsilon}d\delta$. Together, these imply

$$\int_s D_s \max_{o\in O}\left\{b_o T\left(\delta_{so}\right)^{1-\varepsilon}\right\} ds \leq \sum_i \bar{D}_i^k\bar{b}_i^k\int_{s\in\mathcal{S}_i^k}\left|O_i^k\right| G\left(\frac{k^{\mathtt{d}}}{\left|O_i^k\right|}\right) + \bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k T(\delta)^{1-\varepsilon}d\delta$$

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

$$
\begin{aligned}
\pi_j^{\Delta} &= \sup_O \sum_{o\in O} -R_o\xi + Z\left(q_j,|O|\right)^{\varepsilon-1}\int_s D_s \max_{o\in O}\left\{b_o T\left(\delta_{so}\right)^{1-\varepsilon}\right\} ds \\
&\leq \sup_O \sum_{i\in I^k} -\left|O_i^k\right| \underline{R}_i^k \xi + Z\left(q_j,|O|\right)^{\varepsilon-1}\left(\sum_{i\in I^k}\bar{D}_i^k\bar{b}_i^k\left|O_i^k\right| G\left(\frac{k^{\mathtt{d}}}{\left|O_i^k\right|}\right) + \bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k T(\delta)^{1-\varepsilon}d\delta\right) \\
&\leq \sup_{\{N_i\in\mathbb{N}_0\}} \sum_{i\in I^k} -N_i\underline{R}_i^k\xi + Z\left(q_j,\sum_{i'\in I^k}N_{i'}\right)^{\varepsilon-1}\sum_{i\in I^k}\bar{D}_i^k\bar{b}_i^k N_i G\left(\frac{k^{\mathtt{d}}}{N_i}\right) + Z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k T(\delta)^{1-\varepsilon}d\delta \\
&\leq \sup_{\{N_i\geq 0\}} \sum_{i\in I^k} -N_i\underline{R}_i^k\xi + Z\left(q_j,\sum_{i'\in I^k}N_{i'}\right)^{\varepsilon-1}\sum_{i\in I^k}\bar{D}_i^k\bar{b}_i^k N_i G\left(\frac{k^{\mathtt{d}}}{N_i}\right) + Z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k T\left(\delta\right)^{1-\varepsilon}d\delta \\
&= \bar{\pi}_j^{k\Delta}
\end{aligned}
$$

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.12 *Fix k. In the limit as $\Delta\to 0$,*

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

Proof. We begin with the definition of $\bar{\pi}_j^{k\Delta}$, use the expressions for G, T, Z, and ξ in terms of g, t, z, and Δ, and then substitute $n_i \equiv \frac{\Delta^{\mathtt{d}}N_i}{k^{\mathtt{d}}}$ to get

$$
\begin{aligned}
\bar{\pi}_j^{k\Delta} &\equiv \sup_{\{N_i\geq 0\}}\sum_{i\in I^k}\left\{-N_i\underline{R}_i^k\xi + Z\left(q_j,N\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k N_i G\left(\frac{k^{\mathtt{d}}}{N_i}\right)\right\} + Z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k T(\delta)^{1-\varepsilon}d\delta \\
&= \sup_{\{N_i\geq 0\}}\sum_{i\in I^k}\left\{-N_i\underline{R}_i^k\Delta^{\mathtt{d}} + z\left(q_j,\Delta^{\mathtt{d}}N\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k N_i\Delta^{\mathtt{d}} g\left(\frac{1}{\Delta^{\mathtt{d}}}\frac{k^{\mathtt{d}}}{N_i}\right)\right\} + z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}d\delta \\
&= \sup_{\{n_i\geq 0\}} k^{\mathtt{d}}\sum_{i\in I^k}\left\{-n_i\underline{R}_i^k + z\left(q_j,k^{\mathtt{d}}\sum_{\tilde{i}\in I^k}n_{\tilde{i}}\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k n_i g\left(\frac{1}{n_i}\right)\right\} + z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}d\delta
\end{aligned}
$$

Taking the limit gives

$$
\begin{aligned}
\lim_{\Delta\to 0}\bar{\pi}_j^{k\Delta} &= \lim_{\Delta\to 0}\sup_{\{n_i\geq 0\}} k^{\mathtt{d}}\sum_i\left\{-n_i\underline{R}_i^k + z\left(q_j, k^{\mathtt{d}}\sum_{\tilde{i}} n_{\tilde{i}}\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k n_i g\left(\frac{1}{n_i}\right)\right\} + \\
&+ \lim_{\Delta\to 0} z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}d\delta \\
&= \sup_{\{n_i\geq 0\}} k^{\mathtt{d}}\sum_i\left\{-n_i\underline{R}_i^k + z\left(q_j, k^{\mathtt{d}}\sum_{\tilde{i}} n_{\tilde{i}}\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k n_i g\left(\frac{1}{n_i}\right)\right\} \\
&= \sup_{\{n_i\geq 0\}} k^{\mathtt{d}}\sum_i\left\{-n_i\underline{R}_i^k + z\left(q_j, k^{\mathtt{d}}\sum_{\tilde{i}} n_{\tilde{i}}\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k \kappa\left(n_i\right)\right\}
\end{aligned}
$$

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.13 *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 $\kappa(n_s) - \kappa(\bar{n}_j) \leq \kappa'(\bar{n}_j)(n_s - \bar{n}_j) = \frac{\underline{R}}{z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}}(n_s - \bar{n}_j)$. 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.14

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

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 $k < \eta$ 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 $n \in [0, \bar{n}_j]$,

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

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 $\pi_j \leq \bar{\pi}_j^k$ for all k, so taking the limit of both sides yields

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

■

We next bound the payoff from below.

Claim A.15

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

Proof. Begin with

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

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

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

■

Claim A.16 *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 $n_i = \frac{\Delta^{\mathtt{d}} N_i}{k^{\mathtt{d}}}$

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

Next, we use the fact that $\liminf_{\Delta\to 0} \sup_{\{n \geq 0\}} f(n, \Delta) \geq \sup_{\{n \geq 0\}} \liminf_{\Delta \to 0} f(n, \Delta)$[42] along with $\lim_{\Delta\to 0} \Delta^{\mathtt{d}} \left\lceil \frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}} n_i \right\rceil = k^{\mathtt{d}} n_i$ and $\lim_{u\to\infty} \rho(u) = 1$ to get

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

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

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

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

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

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

■

Claim A.17

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

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 Proof of Proposition 2: Convergence of the Policy Function

In this appendix we show uniform convergence of the policy function. We do this in two steps. First, we derive properties of the limiting economy. We show that if the limiting problem has a unique solution, n^*, then for any $\varepsilon > 0$ there exists an $\eta > 0$ such that $n \in \bar{\mathcal{N}}$ and $|\Pi(n) - \Pi(n^*)| < \eta$ imply $\int_{s\in\mathcal{S}} |n_s - n_s^*| \, ds < \varepsilon$, where $\bar{\mathcal{N}}$ is a space of functions with a uniform bound.

In the second step, we study the sequence of economies as $\Delta \to 0$. As in the proof of convergence of the value function in Appendix A.2, we construct a sequence of bounds on the profit function that get tighter as $\Delta \to 0$. We show that for economy Δ, the optimal choice $O^{\Delta*}$ has a corresponding strategy in the limiting economy, $n^{\Delta*}$. As $\Delta \to 0$, the bounds get tighter and two things happen. First, $O^{\Delta*}$ gets close to $n^{\Delta*}$ in the sense that over any Jordan measurable set $\mathcal{A}$, $\Delta^d \left|O^{\Delta*} \cap \mathcal{A}\right|$ uniformly approaches $\int_{s\in\mathcal{A}} n_s^{\Delta*} ds$. Second, the corresponding strategy $n^{\Delta*}$ delivers a value in the limiting economy close to optimum. This, along with the first step, implies that $n^{\Delta*}$ converges to n^*. Namely, we have uniform convergence of the policy function to n^*.

As in the proof of Proposition 1, we use device of using $k \times k$ squares to find upper and lower bounds. In that proof, the key step was to take the limit as $\Delta \to 0$ for a given k and then take $k \to 0$. Here, the key trick is to choose use a sequence of $k = K(\Delta)$, so that as we take the limit as $\Delta \to 0$, the sequence $k = K(\Delta)$ also converges to zero (albeit more slowly than does Δ).

A.3.1 Step 1: Properties of the Limiting Problem

In this section, we show that if there is a unique solution to the limiting problem, then for any $\epsilon > 0$ there exists an $\eta > 0$ such that $|\Pi(n) - \Pi(n^*)| < \eta$ implies $\int_{s\in\mathcal{S}} |n - n_s^*|\, ds < \epsilon$. If the $\bar{\mathcal{N}}$ were compact, then we could use a short, standard proof following Lemma 3.7 of Lucas and Stokey. But we have no reason to believe that $\bar{\mathcal{N}}$ compact. Fortunately, we can use direct methods, and the fact that κ is strictly concave is sufficient.

We beginning by defining several terms.

Let $\mathcal{N} \equiv \{n : \mathcal{S} \to \mathbb{R}\}$ be the set of feasible policies.

Let $\mathcal{N}(N) \equiv \left\{n : \mathcal{S} \to \mathbb{R} \text{ such that } \int_{s\in\mathcal{S}} n_s ds = N\right\}$ be the set of feasible policies for which the firm sets up a measure N of plants.

It will be useful to define a finite upper bound for the measure of plants in a location. Define $\bar{n}$ to satisfy $\kappa'(\bar{n}) = \frac{1}{2}\frac{\underline{R}}{z(q,0)^{\varepsilon-1}\bar{x}}$, where $\bar{x} \equiv (\max_{s\in\mathcal{S}} b_s)(\max_{s\in\mathcal{S}} D_s)$ and $\underline{R} = \min_{s\in\mathcal{S}} R_s$.

Define $\bar{N} \equiv \bar{n}|\mathcal{S}|$ to be an upper bound on the total mass of plants given the upper bounds $\bar{n}$ for any particular location.

Define $\bar{\mathcal{N}} \equiv \{n : \mathcal{S} \to [0, \bar{n}]\}$ to be the set of strategies for which n_s is bounded between 0 and $\bar{n}$, and let $\bar{\mathcal{N}}(N) \equiv \left\{n : \mathcal{S} \to [0, \bar{n}] \text{ such that } \int_{s\in\mathcal{S}} n_s ds = N\right\}$ be the subset of those where the total measure of plants is N.

It will also be useful to have notation for the partial inverse of κ'. $\kappa'(\cdot)$ is strictly decreasing when restricted to a strictly positive domain. Let χ be the partial inverse of κ', so that $\chi(x) \equiv \begin{cases} \kappa'^{-1}(x) & x < \kappa'(0) \\ 0 & x \geq \kappa'(0) \end{cases}$ and note that $\chi(x)$ is continuous.

Lemma A.18 *For any N there exists a unique solution to the problem $\sup_{n\in\mathcal{N}(N)} \Pi(n)$. The optimum $\hat{n}(N)$ and the multiplier $\lambda(N)$ associated with the constraint $\int_{s\in\mathcal{S}} n_s ds = N$ are both continuous in N.*

Proof. Fix N. Consider the problem

$$\max_{n\in\mathcal{N}(N)} \Pi(n) \equiv \max_{n\in\mathcal{N}(N)} \int_{s\in\mathcal{S}} \left[-R_s n_s + x_s z(q,N)^{\varepsilon-1}\kappa(n_s)\right] ds$$

The objective function is strictly concave and the constraint set is convex, so the first order conditions are necessary sufficient to characterize a solution. Letting λ be the multiplier on the constraint $\int_{s\in\mathcal{S}} n_s ds = N$, the first order condition for n_s is

$$R_s + \lambda \geq x_s z^{\varepsilon-1}\kappa'(n_s) \text{ with equality if } n_s > 0.$$

Then the optimal policy and Lagrange multiplier satisfy

$$\begin{aligned}
\hat{n}_s(N) &= \chi\left(\frac{R_s+\lambda\left(N\right)}{x_s z^{\varepsilon-1}}\right) \\
N &= \int_{s\in\mathcal{S}}\chi\left(\frac{R_s+\lambda(N)}{x_s z^{\varepsilon-1}}\right)ds
\end{aligned}$$

Note that for any $N>0$, there is a unique $\lambda(N)$ that satisfies the second equation. The continuity of χ thus implies the continuity $\hat{n}_s(N)$ and $\lambda(N)$ in N. ■

Lemma A.19 *Suppose there is a unique solution, n^*, and that $N^*=\int_{s\in\mathcal{S}}n_s^*ds$. Then*

$$\lim_{N\to N^*}\int_{s\in\mathcal{S}}\left|\hat{n}_s(N)-n_s^*\right|ds=0.$$

Proof. Since n^* is optimal, $\lambda(N^*)=-\frac{d\left[z(q,N^*)^{\varepsilon-1}\right]}{dN^*}\int_{s\in\mathcal{S}}x_s\kappa(n_s^*)ds\geq 0$. $\lambda(N)$ is continuous, so that $\lambda(N)>-\underline{R}/2$ in a neighborhood of N^*. As a result, the first order condition that $R_s+\lambda(N)\geq z(q,N)^{\varepsilon-1}x_s\kappa'(\hat{n}_s(N))$ with equality if $\hat{n}_s(N)>0$ means that we can find a positive lower bound for $\kappa'(\hat{n}_s(N))$ whenever $\lambda(N)>0$. Either $\hat{n}_s(N)=0$ or

$$\kappa'(\hat{n}_s(N))=\frac{R_s+\lambda(N)}{x_s z(q,N)^{\varepsilon-1}}\geq\frac{\underline{R}/2}{\bar{x}z(q,0)^{\varepsilon-1}}.$$

Define $\mathring{n}$ to satisfy $\kappa'(\mathring{n})=\frac{\underline{R}/2}{\bar{x}z(q,0)^{\varepsilon-1}}$. Then $\mathring{n}_s\geq n_s^*$ and, if N is sufficiently close to N^*, $\mathring{n}_s>\hat{n}_s(N)$. Then $|n_s^*-\hat{n}_s(N)|\leq 2\mathring{n}$ when N is close enough to N^*. $2\mathring{n}$ is intregrable over $\mathcal{S}$, and $\hat{n}(N)$ converges to n pointwise, so dominated convergence implies that $\lim_{N\to N^*}\int_{s\in\mathcal{S}}\left|\hat{n}_s(N)-n_s^*\right|ds=0$. ■

Lemma A.20 *The maximization problem $\sup_{n\in\mathcal{N}}\Pi(n)$ obtains a maximum, Π^*. If $n\in\arg\max_{n\in\mathcal{N}}\Pi(n)$ then $\int n_s ds\leq\bar{N}$. If $\int_{s\in\mathcal{S}}n_s ds>\bar{N}$, then $\Pi^*-\Pi(n)\geq\frac{1}{2}\underline{R}(\int_{s\in\mathcal{S}}n_s ds-\bar{N})$.*

Proof. $[0,\bar{N}]$ is a closed, bounded segment of the real line, so it is compact. Further $\Pi\left(\hat{n}(N)\right)$ is continuous in N. Thus $\Pi\left(\hat{n}(N)\right)$ obtains the maximum on $N\in\left[0,\bar{N}\right]$.

We next show that any strategy n such that $\int n_s ds>\bar{N}$ is strictly dominated. Let $\tilde{n}$ be defined so that $\tilde{n}_s=\min\{n_s,\bar{n}\}$. Letting $N\equiv\int_{s\in\mathcal{S}}n_s ds$ and $\tilde{N}\equiv\int_{s\in\mathcal{S}}\tilde{n}_s ds$, note that $z(q,\tilde{N})^{\varepsilon-1}\geq z(q,N)^{\varepsilon-1}$. This implies

$$\begin{aligned}
\Pi\left(\tilde{n}\right)-\Pi(n) &= \int\left\{-R_s\tilde{n}_s+z\left(q,\tilde{N}\right)^{\varepsilon-1}x_s\kappa(\tilde{n}_s)\right\}ds-\int\left\{-R_s n_s+z\left(q,N\right)^{\varepsilon-1}x_s\kappa(n_s)\right\}ds \\
&\geq \int\left\{-R_s\tilde{n}_s+z\left(q,\tilde{N}\right)^{\varepsilon-1}x_s\kappa\left(\tilde{n}_s\right)\right\}ds-\int\left\{-R_s n_s+z\left(q,\tilde{N}\right)^{\varepsilon-1}x_s\kappa\left(n_s\right)\right\}ds \\
&= \int\left\{-R_s\left(\tilde{n}_s-n_s\right)+z\left(q,\tilde{N}\right)^{\varepsilon-1}x_s\left[\kappa(\tilde{n}_s)-\kappa(n_s)\right]\right\}ds
\end{aligned}$$

Since $\kappa(n_s)-\kappa(\tilde{n}_s)\leq\kappa'\left(\bar{n}\right)\left(n_s-\tilde{n}_s\right)=\frac{1}{2}\frac{\underline{R}}{z(q,0)^{\varepsilon-1}\bar{x}}\left(n_s-\tilde{n}_s\right)$, this is

$$\begin{aligned}
\Pi\left(\tilde{n}\right)-\Pi(n) &\geq \int\left\{-R_s\left(\tilde{n}_s-n_s\right)+\frac{1}{2}\frac{z\left(q,\tilde{N}\right)^{\varepsilon-1}}{z\left(q,0\right)^{\varepsilon-1}}\frac{x_s}{\bar{x}}\underline{R}\left(\tilde{n}_s-n_s\right)\right\}ds \\
&= \int\left[\frac{R_s}{\underline{R}}-\frac{1}{2}\frac{z\left(q,\tilde{N}\right)^{\varepsilon-1}}{z\left(q,0\right)^{\varepsilon-1}}\frac{x_s}{\bar{x}}\right]\underline{R}\left(n_s-\tilde{n}_s\right)ds \\
&\geq \int\frac{1}{2}\underline{R}\left(n_s-\tilde{n}_s\right)ds \\
&\geq \int\frac{1}{2}\underline{R}\left(n_s-\bar{n}\right)ds \\
&= \frac{1}{2}\underline{R}\left(\int n_s ds-N\right)
\end{aligned}$$

The conclusion follows from $\int n_s ds>N$ and $Pi^*\geq\Pi(\tilde{n})$. ■

Lemma A.21 *Suppose that there is a unique solution n^*. For any $\gamma>0$, there is an $\eta_1>0$ such that for all $n\in\mathcal{N}$, $|\Pi(n)-\Pi(n^*)|<\eta_1$ implies $\left|\int_{s\in\mathcal{S}}n_s ds-\int_{s\in\mathcal{S}}n_s^* ds\right|<\gamma$.*

Proof. We first consider functions n such that $\int_{s\in\mathcal{S}}n_s ds\in[0,\bar{N}+\gamma]$. Let $N^*=\int_{s\in\mathcal{S}}n_s^* ds$. For $\gamma>0$, define $E_\gamma=\left\{N\in\left[0,\bar{N}+\gamma\right]\text{ such that }|N-N^*|\geq\gamma\right\}$. Lemma A.20 states that $N^*\in[0,\bar{N}]$, so that E_γ is non-empty and compact. For any such γ, let $\eta\equiv\min_{N\in E_\gamma}\left|\max_{n\in\mathcal{N}(N)}\Pi(n)-\Pi(n^*)\right|$. Since the function being minimized is continuous in N and E_γ is compact, the minimum is attained. Moreover, since $N^*\notin E_\gamma$, it follows that $\eta>0$. As a result, any $N\in[0,\bar{N}+\gamma]$ with $|N-N^*|\geq\gamma$ implies that $\left|\max_{n\in\mathcal{N}(N)}\Pi(n)-\Pi(n^*)\right|\geq\eta$.

We next consider functions n is such that $\int_{s\in\mathcal{S}}n_s ds>\bar{N}+\gamma$. For such functions, Lemma A.20 implies that $\Pi(n^*)-\Pi(n)\geq\frac{1}{2}\underline{R}\gamma$.

Together, we have that if $|\Pi(n^*)-\Pi(n)|\leq\eta_1\equiv\min\left\{\eta,\frac{1}{2}\underline{R}\gamma\right\}$ then $\left|\int_{s\in\mathcal{S}}n_s ds-\int_{s\in\mathcal{S}}n_s^* ds\right|<\gamma$.

■

Lemma A.22 *Let $\hat{n}(N)\equiv\arg\max_{n\in\mathcal{N}(N)}\Pi(n)$. For any $\epsilon>0$, there exists a function $h_\epsilon(N)$ that is continuous and strictly positive on $N\in(0,\infty)$, such that for any $\tilde{n}\in\mathcal{N}(N)$, $\Pi\left(\hat{n}(N)\right)-\Pi(\tilde{n})\geq h_\epsilon(N)$ implies $\|\hat{n}(N)-\tilde{n}\|>\epsilon$.*

Proof. Fix $\epsilon>0$ and $N>0$. We will omit the argument N whenever there is no ambiguity.

Let $\lambda(N)$ be the multiplier on the constraint $\int n_s ds=N$ in the maximization problem.

Define $\omega_s(u)\equiv-R_s u-\lambda u+x_s z^{\varepsilon-1}\kappa(u)$. For any $s\in\mathcal{S}$ and $\tau\in(0,\infty)$, $\omega_s(\hat{n}_s)-\omega_s(\hat{n}_s+\tau)$ is strictly positive and strictly increasing in τ. To see this note the fact that $\hat{n}$ is optimal implies that $\omega_s'(\hat{n}_s)\leq 0$ (with equality if $\hat{n}_s>0$). Further, ω_s is strictly concave because it inherits the strict concavity of κ. Thus for any

$\tau > 0$ the strict concavity implies that $\omega_s(\hat{n}_s) - \omega_s(\hat{n}_s + \tau)$ is strictly positive, $\omega_s(\hat{n}_s + \tau) < \omega_s(\hat{n}_s) + \tau\omega_s'(\hat{n}_s) \leq \omega_s(\hat{n}_s)$, and also that it is strictly increasing:

$$\frac{d}{d\tau}\left[\omega_s(\hat{n}_s) - \omega_s(\hat{n}_s + \tau)\right] = -\omega_s'(\hat{n}_s + \tau) > -\omega_s'(\hat{n}_s) \geq 0.$$

Next, define the functions

$$\begin{aligned} H_\epsilon(N) &\equiv \inf_{n \in \mathcal{N}(N)} \Pi\left(\hat{n}(N)\right) - \Pi(n) \text{ subject to } \int_{s \in \mathcal{S}} |n_s - \hat{n}_s|\, ds \geq \epsilon \\ h_\epsilon(N) &\equiv \inf_{\{\tau_s \geq 0\}_{s \in \mathcal{S}}} \int_{s \in \mathcal{S}} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(\hat{n}_s + \tau_s\right)\right] ds \text{ subject to } \int_{s \in \mathcal{S}} \tau_s ds = \frac{\epsilon}{2} \end{aligned}$$

We now show that $H_\epsilon(N) \geq h_\epsilon(N)$. For any n such that $\int_{s \in \mathcal{S}} \hat{n}_s ds = \int_{s \in \mathcal{S}} n_s ds$, we can multiply both sides by $\lambda(N)$ and rearrange to get $\int_{s \in \mathcal{S}} \lambda(N)\left(n_s - \hat{n}_s\right) ds = 0$, meaning that we can rearrange $H_\epsilon(N)$ as

$$H_\epsilon(N) = \inf_{n \in \mathcal{N}(N)} \int_{s \in \mathcal{S}} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(n_s\right)\right] ds \text{ subject to } \int_{s \in \mathcal{S}} |n_s - \hat{n}_s|\, ds \geq \epsilon$$

We can rearrange this further as

$$H_\epsilon(N) = \inf_{n \in \mathcal{N}(N)} \int_{s \in \mathcal{S} | n_s < \hat{n}_s} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(n_s\right)\right] ds + \int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(n_s\right)\right] ds$$

subject to $\int_{s \in \mathcal{S} | n_s < \hat{n}_s} |n_s - \hat{n}_s|\, ds = \int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} |n_s - \hat{n}_s|\, ds$ and $\int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} |n_s - \hat{n}_s|\, ds \geq \frac{\epsilon}{2}$. Since $\omega_s\left(\hat{n}_s\right) \geq \omega_s\left(n_s\right)$,we have

$$H_\epsilon(N) \geq \inf_{n \in \mathcal{N}(N)} \int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(n_s\right)\right] ds$$

subject to $\int_{s \in \mathcal{S} | n_s < \hat{n}_s} |n_s - \hat{n}_s|\, ds = \int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} |n_s - \hat{n}_s|\, ds$ and $\int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} |n_s - \hat{n}_s|\, ds \geq \frac{\epsilon}{2}$. Relaxing a constraint gives a weakly smaller number, so that

$$\begin{aligned} H_\epsilon(N) &\geq \inf_{n \in \mathcal{N}(N)} \int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(n_s\right)\right] ds \text{ subject to } \int_{s \in \mathcal{S} | n_s \geq \hat{n}_s} |n_s - \hat{n}_s|\, ds \geq \frac{\epsilon}{2} \\ &\geq \inf_{\{\tau_s \geq 0\}_{s \in \mathcal{S}}} \int_{s \in \mathcal{S}} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(\hat{n}_s + \tau_s\right)\right] ds \text{ subject to } \int_{s \in \mathcal{S}} \tau_s ds \geq \frac{\epsilon}{2} \end{aligned}$$

Since $\omega_s\left(\hat{n}_s\right) - \omega_s\left(\hat{n}_s + \tau\right)$ is increasing and convex in τ on $\tau \in [0, \infty)$, the value of the right hand side is unchanged if we impose that the constraint holds with equality. This gives

$$\begin{aligned} H_\epsilon(N) &\geq \inf_{\{\tau_s \geq 0\}_{s \in \mathcal{S}}} \int_{s \in \mathcal{S}} \left[\omega_s\left(\hat{n}_s\right) - \omega_s\left(\hat{n}_s + \tau_s\right)\right] ds \text{ subject to } \int_{s \in \mathcal{S}} \tau_s ds = \frac{\epsilon}{2} \\ &= h_\epsilon(N) \end{aligned}$$

We next show that $h_\epsilon(N)$ is strictly increasing in ϵ. To do this, we solve the minimization problem. The objective function is strictly convex and the constraint set is convex, so the first order conditions are necessary and sufficient for a minimum. Let μ be the multiplier on the constraint. Then the first order condition for τ_s is

$$\omega_s'\left(\hat{n}_s+\tau_s\right)+\mu \leq 0 \text{ with equality if } \tau_s>0$$

or, equivalently,

$$-R_s-\lambda(N)+x_s z^{\varepsilon-1}\kappa'\left(\hat{n}_s+\tau_s\right)+\mu \leq 0 \quad \text{with equality if} \quad \tau_s>0$$

Since ω' is continuous, τ_s is continuous is s.

Note that it must be that $\mu>0$: otherwise the FOC for τ_s would imply $\tau_s=0$, $\forall s$ because $\omega_s\left(\hat{n}_s+\tau\right)$ is strictly concave on $\tau \in (0,\infty)$ and $\omega_s'\left(\hat{n}_s\right) \leq 0$, $\forall s$, and this would violate the constraint. Note also that $R_s+\lambda>\mu$ for all s: $\lim_{\tau\to\infty}\omega_s\left(n+\tau\right)=-\infty$ implies that the optimal τ is finite, so that the FOC implies $-R_s-\lambda(N)+\mu \leq -x_s z^{\varepsilon-1}\kappa'\left(\hat{n}_s+\tau_s\right)<0$:. Recalling that χ is the partial inverse of κ' and that χ is continuous, the first order condition for τ_s can be restated as

$$\tau_s=\max\left\{0,\chi\left(\frac{R_s+\lambda-\mu}{x_s z^{\varepsilon-1}}\right)-\hat{n}_s\right\}. \tag{18}$$

μ must therefore satisfy

$$\int_{s\in\mathcal{S}}\max\left\{0,\chi\left(\frac{R_s+\lambda-\mu}{x_s z^{\varepsilon-1}}\right)-\hat{n}_s\right\}=\frac{\epsilon}{2}$$

The continuity and monotonicity of χ implies that this has a unique solution, $\mu(\epsilon,N)$ that is continuous in ϵ. Note also that $\mu(\epsilon,N)$ is continuous in N for fixed ϵ.

We next show that $h_\epsilon(N)$ is strictly increasing in ϵ. Consider $\epsilon>0$. Let $\tau_s(\epsilon)$ be the solution for ϵ described by (18). Note that τ_s is continuous in s. There must be an $\bar{\eta}>0$ such that the set $E=\{s|\tau_s\left(\epsilon\right)\geq\bar{\eta}\}$ has strictly positive measure (if no such $\bar{\eta}$, E existed, it would violate the constraint $\int_{s\in\mathcal{S}}\tau_s(\epsilon)ds\geq\frac{\epsilon}{2}$). Since τ_s is continuous and E is bounded, E is compact. Consider ϵ' such that $0\leq\frac{\epsilon}{2}-\bar{\eta}|E|<\frac{\epsilon'}{2}<\frac{\epsilon}{2}$. For any such ϵ', let $\eta=\frac{1}{|E|}\left(\frac{\epsilon}{2}-\frac{\epsilon'}{2}\right)\in(0,\bar{\eta})$. Consider the strategy of $\tau_s(\epsilon')=\tau_s(\epsilon)-\eta$ if $s\in E$ and $\tau_s(\epsilon')=\tau_s(\epsilon)$ if $s\notin E$. Then

$$\begin{aligned}
h_{\epsilon'}(N)-h_\epsilon(N) &\leq \int_{s\in\mathcal{S}}\left[\omega_s\left(\hat{n}_s+\tau_s\left(\epsilon\right)\right)-\omega_s\left(\hat{n}_s+\tau_s\left(\epsilon'\right)\right)\right]ds\\
&= \int_{s\in E}\left[\omega_s\left(\hat{n}_s+\tau_s\left(\epsilon\right)\right)-\omega_s\left(\hat{n}_s+\tau_s\left(\epsilon\right)-\eta\right)\right]ds\\
&\leq |E|\sup_{s\in E}\left[\omega_s\left(\hat{n}_s+\tau_s\left(\epsilon\right)\right)-\omega_s\left(\hat{n}_s+\tau_s\left(\epsilon\right)-\eta\right)\right]
\end{aligned}$$

Since E is compact while $\omega_s\left(\hat{n}_s+\tau_s(\epsilon)\right)-\omega_s\left(\hat{n}_s+\tau_s(\epsilon)-\eta\right)$ is continuous and strictly negative, the supremum is strictly negative, giving $h_{\epsilon'}(N)<h_\epsilon(N)$.

We next establish that for $\epsilon > 0$, $h_\epsilon(N)$ is strictly positive and continuous in N. The fact that $h_\epsilon(N)$ is strictly increasing in ϵ and $h_0(N) = 0$ implies that $h_\epsilon(N)$ is strictly positive when $\epsilon > 0$. Further, continuity in N follows from the continuity of $\hat{n}(N)$, $\tau_s(\epsilon, N)$, $z(q, N)$, $\lambda(N)$, and $\mu(\epsilon, N)$ in N and the continuity of $\omega_s(\hat{n}_s) - \omega_s(\hat{n}_s + \tau)$ in τ. ■

Claim A.23 *Suppose that there is a unique solution, n^*. Then for any $\epsilon > 0$, there is an $\eta > 0$ such that for any $n \in \mathcal{N}$, $|\Pi(n) - \Pi(n^*)| < \eta$ implies $\int_{s\in\mathcal{S}} |n_s - n_s^*|\, ds < \epsilon$.*

Proof. Fix $\epsilon > 0$. Let $N^* \equiv \int_{s\in\mathcal{S}} n_s^* ds$ and let $\hat{n}(N) \equiv \arg\max_{n\in\mathcal{N}(N)} \Pi(n)$. Using Lemma A.19, there is a $\gamma > 0$ such that $|N - N^*| < \gamma$ implies that $\int_{s\in\mathcal{S}} |n_s^* - \hat{n}_s(N)|\, ds < \frac{\epsilon}{2}$. Define

$$J(\epsilon, N) = \inf_{n\in\mathcal{N}(N)} \Pi(n^*) - \Pi(n) \text{ subject to } \int_{s\in\mathcal{S}} |n_s - n_s^*|\, ds \geq \epsilon$$

We will show that there is a strictly-positive, uniform lower bound $\underline{J}(\epsilon)$ on $J(\epsilon, N)$ for any $N \in [N^* - \gamma, N^* + \gamma]$.

Since $\Pi(n^*) \geq \Pi(\hat{n}(N))$, we have

$$J(\epsilon, N) \geq \inf_{n\in\mathcal{N}(N)} \Pi(\hat{n}(N)) - \Pi(n) \text{ subject to } \int_{s\in\mathcal{S}} |n_s - n_s^*| ds \geq \epsilon$$

Relaxing the constraint delivers a lower bound. In particular, since $\int_{s\in\mathcal{S}} |n_s - n_s^*|\, ds \leq \int_{s\in\mathcal{S}} |n_s - \hat{n}_s(N)|\, ds + \int_{s\in\mathcal{S}} |\hat{n}_s(N) - n_s^*|\, ds$, the constraint can be relaxed to $\int_{s\in\mathcal{S}} |n_s - \hat{n}_s(N)|\, ds \geq \epsilon - \int_{s\in\mathcal{S}} |\hat{n}_s(N) - n_s^*|\, ds$, and since $\int_{s\in\mathcal{S}} |\hat{n}_s(N) - n_s^*|\, ds < \frac{\epsilon}{2}$, a further relaxation is $\int_{s\in\mathcal{S}} |n_s - \hat{n}_s(N)|\, ds \geq \frac{\epsilon}{2}$. As a result,

$$J(\epsilon, N) \geq \inf_{n\in\mathcal{N}(N)} \Pi(\hat{n}(N)) - \Pi(n_s) \text{ subject to } \int_{s\in\mathcal{S}} |n_s - \hat{n}_s(N)|\, ds \geq \frac{\epsilon}{2}$$

Using the previous lemma, this implies that $J(\epsilon, N) \geq h_{\epsilon/2}(N)$. Define the uniform lower bound

$$\begin{aligned} \underline{J}(\epsilon) &\equiv \inf_{N\in[N^*-\gamma, N^*+\gamma]} J(\epsilon, N) \\ &\geq \inf_{N\in[N^*-\gamma, N^*+\gamma]} h_{\epsilon/2}(N) \end{aligned}$$

Since $[N^* - \gamma, N^* + \gamma]$ is compact and $h_{\epsilon/2}(N)$ is continuous and strictly positive, the infimum achieves a minimum which is strictly positive, i.e., $\underline{J}(\epsilon) > 0$, which further implies that $J(\epsilon, N) \geq \underline{J}(\epsilon) > 0$, $\forall N \in [N^* - \gamma, N^* + \gamma]$. To summarize, we have established that if $\Pi(n^*) - \Pi(n) < \bar{J}(\epsilon)$ and $\left|\int n_s ds - N^*\right| < \gamma$ then $\int_{s\in S} |n - n_s^*|\, ds < \epsilon$.

According to Lemma A.21, there is an $\eta_1 > 0$ such that $|\Pi(n) - \Pi(n^*)| < \eta_1$ implies that $\left|\int_{s\in\mathcal{S}} n_s ds - N^*\right| < \gamma$. Together, these two results imply that if

$$|\Pi(n) - \Pi(n^*)| < \eta \equiv \min\{\eta_1, \underline{J}(\epsilon)\}$$

then $\int_{s\in S}|n-n_s^*|\,ds<\epsilon$. ■

A.3.2 Step 2: Convergence of the policy function

In this section, we show that if there is a unique solution to the limiting problem, then the appropriately scaled policy function converges uniformly as $\Delta\to 0$.

As in the proof of the Proposition 1, we use the line segments of length k or $k\times k$ squares to derive bounds. In Proposition 1, the key step was to take the limit as $\Delta\to 0$ for a given k and then take $k\to 0$. We cannot use that strategy here, because, if k is held fixed, as we take a limit $\Delta\to 0$, neither the sequence of strategies for the upper bound nor the sequence of strategies for the lower bound converge to n^*. Instead, the key trick here is, as we take $\Delta\to 0$, to let k approach zero as well. Specifically, we define a function the sequence $K(\Delta)$ converges to zero, but more slowly than does Δ in a sense described below. The fact that k approaches zero as well ensures uniform convergence of the (appropriately scaled) policy function.

Let I^k be the set of line segments of length k or squares of size $k\times k$, so that $\mathcal{S}=\bigcap_{i\in I^k}\mathcal{S}_i^k$.

Let $\Pi^\Delta(O)$ be the profit a firm would get in economy Δ if it chooses a set of plants O. For a vector $\mathbf{N}=\{N_i\}_{i\in I^k}$, define

$$U^{k,\Delta}(\mathbf{N})\equiv\sup_O\Pi^\Delta(O)\ \text{ subject to }\ \left|O\cap\mathcal{S}_i^k\right|=N_i$$

to be profit in economy Δ when the policy is constrained so that the firm places N_i plants in $\mathcal{S}_i^k$, $\forall i\in I_k$. In addition, define

$$\hat{n}^{k,\Delta}(\mathbf{N})\equiv\arg\max_n\Pi(n)\qquad\text{subject to}\quad\int_{s\in\mathcal{S}_i^k}n_s ds=\Delta^{\mathtt{d}}N_i$$

to be the optimal policy in the limiting economy under the constraints that a measure $\Delta^{\mathtt{d}}N_i$ of plants is placed in $\mathcal{S}_i^k$.

For any k, Δ, and any $\mathbf{N}=\{N_i\}_{i\in I^k}$, define $\bar{U}^{k,\Delta}(\mathbf{N})$ and $\underline{U}^{k,\Delta}(\mathbf{N})$ to be upper and lower bounds on the profit a firm could achieve in economy Δ if it chose to place N_i plants in $\mathcal{S}_i^k$:

$$\begin{aligned}\bar{U}^{k,\Delta}(\mathbf{N})&\equiv\sum_{i\in I^k}-\underline{R}_i^k\Delta^{\mathtt{d}}N_i+z\left(q,\Delta^{\mathtt{d}}\sum_{\tilde{\imath}\in I^k}N_{\tilde{\imath}}\right)^{\varepsilon-1}\sum_{i\in I^k}\bar{D}_i^k\bar{b}_i^kN_i\Delta^{\mathtt{d}}g\left(\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}N_i}\right)+z(q,0)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}-3}\int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}dk\\ \underline{U}^{k,\Delta}(\mathbf{N})&\equiv\sum_{i\in I^k}-\bar{R}_i^k\Delta^{\mathtt{d}}N_i+z\left(q,\Delta^{\mathtt{d}}\sum_{\tilde{\imath}\in I^k}N_{\tilde{\imath}}\right)^{\varepsilon-1}\sum_{i\in I^k}\underline{D}_i^k\underline{b}_i^kN_i\Delta^{\mathtt{d}}g\left(\rho(N_i)\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}N_i}\right)\end{aligned}$$

It follows from the same arguments of Claims A.11 and A.15 that

$$\underline{U}^{k,\Delta}(\mathbf{N})\le U^{k,\Delta}(\mathbf{N})\le\bar{U}^{k,\Delta}(\mathbf{N})$$

Define $K(\Delta)$ to be an increasing function that is increasing more slowly than Δ, so that $\lim_{\Delta\to 0}K(\Delta)=0$

but $\lim_{\Delta\to 0} K(\Delta)^{\mathtt{d}-3}\Delta = 0$, and $\frac{1}{K(\Delta)}$ is an integer, e.g., $K(\Delta) = \lceil \Delta^{-1/3} \rceil^{-1}$. One implication is that $\lim_{\Delta\to 0} \frac{\Delta}{K(\Delta)} = \lim_{\Delta\to 0} K(\Delta)^{2-\mathtt{d}} K(\Delta)^{\mathtt{d}-3}\Delta = 0$.

Lemma A.24 *Fix $\epsilon > 0$. There exists a $\bar{\Delta} > 0$ such that for any $\Delta < \bar{\Delta}$ and any $\mathbf{N} = \{N_i\}_{i\in I^{K(\Delta)}}$ such that $\sum_{i\in I^{K(\Delta)}} N_i \leq \frac{2\bar{N}}{\Delta^{\mathtt{d}}}$,*

$$\left|\bar{U}^{K(\Delta),\Delta}(\mathbf{N}) - \underline{U}^{K(\Delta),\Delta}(\mathbf{N})\right| < \epsilon$$

Proof. Fix $\mathbf{N}$, and let $n_i = \frac{\Delta^{\mathtt{d}}}{k^{\mathtt{d}}} N_i$. Then we have

$$\begin{aligned}
\left|\bar{U}^{k,\Delta}(\mathbf{N}) - \underline{U}^{k,\Delta}(\mathbf{N})\right| \leq\ & k^{\mathtt{d}} \sum_i \left\{ -n_i \underline{R}_i^k + z\left(q_j, k^{\mathtt{d}} \sum_i n_i\right)^{\varepsilon-1} \bar{D}_i^k \bar{b}_i^k n_i g\left(\frac{1}{n_i}\right)\right\} \\
& + z\left(q_j, 0\right)^{\varepsilon-1} \bar{D}\bar{b} k^{\mathtt{d}-3} \int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon} d\delta \\
& - k^{\mathtt{d}} \sum_i \left\{ -n_i \bar{R}_i^k + z\left(q_j, k^{\mathtt{d}} \sum_i n_i\right)^{\varepsilon-1} \underline{D}_i^k \underline{b}_i^k n_i g\left(\rho\left(\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}} n_i\right)\frac{1}{n_i}\right)\right\}
\end{aligned}$$

This can be rearranged as

$$\left|\bar{U}^{k,\Delta}(\mathbf{N}) - \underline{U}^{k,\Delta}(\mathbf{N})\right| = A_1^{k,\Delta}(\mathbf{N}) + A_2^{k,\Delta}(\mathbf{N}) + A_3^{k,\Delta}(\mathbf{N})$$

where

$$\begin{aligned}
A_1^{k,\Delta}(\mathbf{N}) &\equiv z(q_j, 0)^{\varepsilon-1} \bar{D}\bar{b} k^{\mathtt{d}-3} \int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon} d\delta \\
A_2^{k,\Delta}(\mathbf{N}) &\equiv k^{\mathtt{d}} \sum_i \left\{ -n_i\left(\underline{R}_i^k - \bar{R}_i^k\right) + z\left(q_j, k^{\mathtt{d}} \sum_{\tilde{\imath}\in I^k} n_{\tilde{\imath}}\right)^{\varepsilon-1} \left(\bar{D}_i^k \bar{b}_i^k - \underline{D}_i^k \underline{b}_i^k\right) n_i g\left(\frac{1}{n_i}\right)\right\} \\
A_3^{k,\Delta}(\mathbf{N}) &\equiv k^{\mathtt{d}} \sum_i \left\{ z\left(q_j, k^{\mathtt{d}} \sum_i n_i\right)^{\varepsilon-1} \underline{D}_i^k \underline{b}_i^k n_i \left[g\left(\frac{1}{n_i}\right) - g\left(\rho\left(\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}} n_i\right)\frac{1}{n_i}\right)\right]\right\}
\end{aligned}$$

We bound each of these three terms separately. First, note that since $\lim_{\Delta\to 0} K(\Delta)^{\mathtt{d}-3}\Delta = 0$, there is a $\bar{\Delta}_1$ small enough so that $\Delta < \bar{\Delta}_1$ implies that $K(\Delta)^{\mathtt{d}-3}\Delta < \frac{1}{z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}\int_0^\infty t(\tilde{\delta})^{1-\varepsilon} d\tilde{\delta}} \frac{\epsilon}{3}$. This means that for

$\Delta < \bar{\Delta}_1$:

$$\begin{aligned} A_1^{K(\Delta),\Delta}(\mathbf{N}) &= z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}K(\Delta)^{\mathtt{d}-3}\int_0^{K(\Delta)} t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon} d\delta \\ &= z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}K(\Delta)^{\mathtt{d}-3}\Delta\int_0^{\frac{K(\Delta)}{\Delta}} t\left(\tilde{\delta}\right)^{1-\varepsilon} d\tilde{\delta} \\ &< \left(z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}\int_0^{\frac{K(\Delta)}{\Delta}} t\left(\tilde{\delta}\right)^{1-\varepsilon} d\tilde{\delta}\right)\frac{1}{z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}\int_0^\infty t\left(\tilde{\delta}\right)^{1-\varepsilon} d\tilde{\delta}}\frac{\epsilon}{3} \\ &= \frac{\int_0^{\frac{K(\Delta)}{\Delta}} t\left(\tilde{\delta}\right)^{1-\varepsilon} d\tilde{\delta}}{\int_0^\infty t\left(\tilde{\delta}\right)^{1-\varepsilon} d\tilde{\delta}}\frac{\epsilon}{3} \\ &\le \frac{\epsilon}{3} \end{aligned}$$

where the second line used the change of variables $\tilde{\delta} = \frac{\delta}{\Delta}$.

We turn next to the second term, $A_2^{k,\Delta}(\mathbf{N})$. Since R_s, D_s, and b_s are uniformly continuous, there is a $\bar{k}$ such that $k < \bar{k}$ implies $\bar{R}_i^k - \underline{R}_i^k \le \frac{\epsilon}{6(2\bar{N})}$ and $\bar{D}_i^k\bar{b}_i^k - \underline{D}_i^k\underline{b}_i^k < \frac{1}{z(q_j,0)^{\varepsilon-1}|\mathcal{S}|}\frac{\epsilon}{6}$. Thus $k < \bar{k}$ implies

$$\begin{aligned} A_2^{k,\Delta}(\mathbf{N}) &= k^{\mathtt{d}}\sum_i\left\{-n_i\left(\underline{R}_i^k - \bar{R}_i^k\right) + z\left(q_j, k^{\mathtt{d}}\sum_i n_i\right)^{\varepsilon-1}\left(\bar{D}_i^k\bar{b}_i^k - \underline{D}_i^k\underline{b}_i^k\right)n_i g\left(\frac{1}{n_i}\right)\right\} \\ &\le k^{\mathtt{d}}\sum_i\left\{n_i\frac{\epsilon}{6\left(2\bar{N}\right)} + \frac{z\left(q_j, k^{\mathtt{d}}\sum_i n_i\right)^{\varepsilon-1}}{z\left(q_j,0\right)^{\varepsilon-1}|\mathcal{S}|}\frac{\epsilon}{6}n_i g\left(\frac{1}{n_i}\right)\right\} \\ &\le \frac{\epsilon}{6} + \frac{\epsilon}{6} \\ &= \frac{\epsilon}{3} \end{aligned}$$

where the third line follows from $k^{\mathtt{d}}\sum_i n_i = \Delta^{\mathtt{d}}\sum_i N_i \le 2\bar{N}$, $z\left(q_j, k^{\mathtt{d}}\sum_i n_i\right) \le z\left(q_j, 0\right)$, and $n_i g\left(\frac{1}{n_i}\right) \le 1$. If $\bar{\Delta}_2$ is such that $\Delta < \bar{\Delta}_2$ implies that $K\left(\Delta\right) < \bar{k}$, $\Delta < \bar{\Delta}_2$ implies that $A_2^{K(\Delta),\Delta}\left(\mathbf{N}\right) < \frac{\epsilon}{3}$.

We turn next to the third term. Using $z\left(q_j, k^{\mathtt{d}}\sum_i n_i\right)^{\varepsilon-1}\underline{D}_i^k\underline{b}_i^k \le z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}$ along with $g\left(\frac{1}{n_i}\right) \ge g\left(\rho\left(\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}}n_i\right)\frac{1}{n_i}\right)$ (which follows from the fact that g is increasing and $\rho\left(\cdot\right) \le 1$), we have

$$\begin{aligned} A_3^{k,\Delta}(\mathbf{N}) &= k^{\mathtt{d}}\sum_i\left\{z\left(q_j, k^{\mathtt{d}}\sum_i n_i\right)^{\varepsilon-1}\underline{D}_i^k\underline{b}_i^k n_i\left[g\left(\frac{1}{n_i}\right) - g\left(\rho\left(\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}}n_i\right)\frac{1}{n_i}\right)\right]\right\} \\ &\le z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}k^{\mathtt{d}}\sum_i n_i\left[g\left(\frac{1}{n_i}\right) - g\left(\rho\left(\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}}n_i\right)\frac{1}{n_i}\right)\right] \end{aligned}$$

Let $u > 0$ be small enough so that such that $ug\left(\frac{1}{u}\right) < \frac{1}{z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}}\frac{\epsilon}{3}$. Since $ng\left(\frac{1}{n}\right)$ is increasing, we have for any $n_i \leq u$ and any Δ,

$$n_i\left[g\left(\frac{1}{n_i}\right) - g\left(\rho\left(\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}}n_i\right)\frac{1}{n_i}\right)\right] \leq n_i g\left(\frac{1}{n_i}\right) \leq \frac{1}{z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}}\frac{\epsilon}{3}$$

Let $\bar{\Delta}_3$ be such that $\Delta < \bar{\Delta}_3$ implies $1-\rho\left(\frac{K(\Delta)^{\mathtt{d}}}{\Delta^{\mathtt{d}}}u\right) \leq \frac{1}{z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}g'(0)}\frac{\epsilon}{3}$. Such a $\bar{\Delta}_3$ exists because $\lim_{x\to\infty}\rho(x) = 1$. For i such that $n_i > u$, we can then bound the term using the fact that g is concave, which implies $g\left(\frac{1}{n}\right) \leq g\left(\rho\left(\cdot\right)\frac{1}{n}\right) + g'\left(\rho\left(\cdot\right)\frac{1}{n}\right)\frac{1}{n_i}\left[1-\rho\left(\cdot\right)\right] \leq g\left(\rho\left(\cdot\right)\frac{1}{n}\right) + g'(0)\frac{1}{n_i}\left[1-\rho\left(\cdot\right)\right]$. Together, these imply for $n_i > u$ and $\Delta < \bar{\Delta}_3$

$$\begin{aligned}
n_i\left[g\left(\frac{1}{n_i}\right) - g\left(\rho\left(\frac{K\left(\Delta\right)^{\mathtt{d}}}{\Delta^{\mathtt{d}}}n_i\right)\frac{1}{n_i}\right)\right] &\leq g'(0)\left[1-\rho\left(\frac{K\left(\Delta\right)^{\mathtt{d}}}{\Delta^{\mathtt{d}}}n_i\right)\right] \\
&\leq g'(0)\left[1-\rho\left(\frac{K\left(\Delta\right)^{\mathtt{d}}}{\Delta^{\mathtt{d}}}u\right)\right] \\
&\leq g'(0)\frac{1}{z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}g'(0)}\frac{\epsilon}{3}
\end{aligned}$$

where the second inequality used the fact that ρ is increasing and $n_i \geq u$. Together, these imply that for any n_i and $\Delta < \bar{\Delta}_3$,

$$n_i\left[g\left(\frac{1}{n_i}\right) - g\left(\rho\left(\frac{K\left(\Delta\right)^{\mathtt{d}}}{\Delta^{\mathtt{d}}}n_i\right)\frac{1}{n_i}\right)\right] \leq \frac{1}{z\left(q_j,0\right)^{\varepsilon-1}\bar{D}\bar{b}}\frac{\epsilon}{3}$$

As a result, if $\Delta < \bar{\Delta}_3$, then $A_3^{K(\Delta),\Delta}\left(\mathbf{N}\right) \leq \frac{\epsilon}{3}$.

These three results together imply that if $\Delta < \min\left\{\bar{\Delta}_1, \bar{\Delta}_2, \bar{\Delta}_3\right\}$, then

$$\left|\bar{U}^{k,\Delta}\left(\mathbf{N}\right) - \underline{U}^{k,\Delta}\left(\mathbf{N}\right)\right| < \epsilon$$

■

Lemma A.25 *For any $m > 0$, there is an $\bar{\Delta} > 0$ such that $\Delta < \bar{\Delta}$ implies $\Delta^{\mathtt{d}}\left|O^{\Delta *}\right| \leq \bar{N} + m$ for any optimal choice $O^{\Delta *}$.*

Proof. Fix $m > 0$. For any k, Δ, consider a vector $\mathbf{N} = \{N_i\}_{i\in I^k}$, along with the alternative vector $\tilde{\mathbf{N}} = \left\{\tilde{N}_i\right\}_{i\in I^k}$ where $\tilde{N}_i = \min\left\{N_i, \left\lceil\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}}\bar{n}\right\rceil\right\}$. We first derive an upper bound on $\bar{U}^{k,\Delta}\left(\mathbf{N}\right) - \bar{U}^{k,\Delta}\left(\tilde{\mathbf{N}}\right)$. Define $n_i = \frac{\Delta^{\mathtt{d}}}{k^{\mathtt{d}}}N_i$ and $\tilde{n}_i = \frac{\Delta^{\mathtt{d}}}{k^{\mathtt{d}}}\tilde{N}_i$. Define $a^{k,\Delta} \equiv \frac{\Delta^{\mathtt{d}}}{\bar{n}k^{\mathtt{d}}}\left\lceil\frac{k^{\mathtt{d}}}{\Delta^{\mathtt{d}}}\bar{n}\right\rceil$, so that $\tilde{n}_i = \min\left\{n_i, a^{k,\Delta}\bar{n}\right\}$. There is a $\bar{\Delta}_1$ such that $\Delta < \bar{\Delta}_1$ implies $a^{K(\Delta),\Delta} < 1 + \frac{m}{3\bar{N}}$.

Noting that $z\left(q_j, k^{\mathtt{d}}\sum_i \tilde{n}_i\right) \geq z\left(q_j, k^{\mathtt{d}}\sum_i n_i\right)$, (and using the shorthand $\tilde{z} \equiv z\left(q_j, k^{\mathtt{d}}\sum_i \tilde{n}_i\right)$), we have:

$$\begin{aligned}
\bar{U}^{k,\Delta}(\mathbf{N}) - \bar{U}^{k,\Delta}\left(\tilde{\mathbf{N}}\right) &= k^{\mathtt{d}}\sum_{i\in I^k}\left\{-n_i\underline{R}_i^k + z\left(q_j, k^{\mathtt{d}}\sum_i n_i\right)^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k\kappa(n_i)\right\} \\
&\quad + z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}\frac{1}{k}\int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}d\delta - \bar{U}^{k,\Delta}\left(\tilde{\mathbf{N}}\right) \\
&\leq k^{\mathtt{d}}\sum_{i\in I^k}\left\{-n_i\underline{R}_i^k + \tilde{z}^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k\kappa(n_i)\right\} + z(q_j,0)^{\varepsilon-1}\bar{D}\bar{b}\frac{1}{k}\int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon}d\delta - \bar{U}^{k,\Delta}\left(\tilde{\mathbf{N}}\right) \\
&= k^{\mathtt{d}}\sum_{i\in I^k}\left\{-n_i\underline{R}_i^k + \tilde{z}^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k\kappa(n_i)\right\} - k^{\mathtt{d}}\sum_{i\in I^{k+}}\left\{-\tilde{n}_i\underline{R}_i^k + \tilde{z}^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k\kappa(\tilde{n}_i)\right\} \\
&= k^{\mathtt{d}}\sum_{i\in I^{k+}}\left\{-\left(n_i - a^{k,\Delta}\bar{n}\right)\underline{R}_i^k + \tilde{z}^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k\left[\kappa(n_i) - \kappa\left(a^{k,\Delta}\bar{n}\right)\right]\right\}
\end{aligned}$$

where I^{k+} is the set of squares such that $N_i > \tilde{N}_i$ (and hence $n_i > \tilde{n}_i = a^{k,\Delta}\bar{n}$). The concavity of κ and $a^{k,\Delta} \geq 1$ implies that $\kappa(n_i) - \kappa\left(a^{k,\Delta}\bar{n}\right) \leq \kappa'\left(a^{k,\Delta}\bar{n}\right)\left(n_i - a^{k,\Delta}\bar{n}\right) \leq \kappa'(\bar{n})\left(n_i - a^{k,\Delta}\bar{n}\right) = \frac{1}{2}\frac{\underline{R}}{z(q,0)^{\varepsilon-1}\bar{D}\bar{b}}\left(n_i - a^{k,\Delta}\bar{n}\right)$. This gives

$$\begin{aligned}
\bar{U}^{k,\Delta}(\mathbf{N}) - \bar{U}^{k,\Delta}\left(\tilde{\mathbf{N}}\right) &\leq k^{\mathtt{d}}\sum_{i\in I^{k+}}\left\{-\left(n_i - a^{k,\Delta}\bar{n}\right)\underline{R}_i^k + \frac{1}{2}\tilde{z}^{\varepsilon-1}\bar{D}_i^k\bar{b}_i^k\frac{\underline{R}}{z(q,0)^{\varepsilon-1}\bar{D}\bar{b}}\left(n_i - a^{k,\Delta}\bar{n}\right)\right\} \\
&\leq k^{\mathtt{d}}\sum_{i\in I^{k+}}\left[-1+\frac{1}{2}\right]\underline{R}\left(n_i - a^{k,\Delta}\bar{n}\right) \\
&= -\frac{\underline{R}}{2}k^{\mathtt{d}}\sum_{i\in I^{k+}}\left(n_i - a^{k,\Delta}\bar{n}\right)
\end{aligned}$$

In particular, if $\Delta^{\mathtt{d}}\sum_{i\in I^k} N_i \geq \bar{N} + m$ and $\Delta < \bar{\Delta}_1$, then

$$\begin{aligned}
K(\Delta)^{\mathtt{d}}\sum_{i\in I^{K(\Delta)+}}\left(n_i - a^{K(\Delta),\Delta}\bar{n}\right) &\geq K(\Delta)^{\mathtt{d}}\sum_{i\in I^{K(\Delta)}}\left(n_i - a^{K(\Delta),\Delta}\bar{n}\right) \\
&= K(\Delta)^{\mathtt{d}}\sum_{i\in I^{K(\Delta)}} n_i - a^{K(\Delta),\Delta}\bar{N} \\
&= \Delta^{\mathtt{d}}\sum_{i\in I^{K(\Delta)}} N_i - a^{K(\Delta),\Delta}\bar{N} \\
&\geq \bar{N} + m - a^{K(\Delta),\Delta}\bar{N} \\
&\geq \bar{N} + m - \left(1+\frac{m}{3\bar{N}}\right)\bar{N} \\
&\geq \frac{2}{3}m
\end{aligned}$$

which would imply

$$\bar{U}^{K(\Delta),\Delta}(\mathbf{N}) - \bar{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}\right) \leq -\frac{R}{3}m$$

Let $\bar{\Delta}_2 > 0$ be such that for any $\Delta < \bar{\Delta}_2$ and any $\tilde{\mathbf{N}} = \left\{\tilde{N}_i\right\}_{i \in I^{K(\Delta)}}$ such that $\sum_{i \in I^{K(\Delta)}} \tilde{N}_i \leq \frac{2\bar{N}}{\Delta^{\mathsf{d}}}$,

$$\left|\bar{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}\right) - \underline{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}\right)\right| < \frac{R}{4}m$$

. The existence of Such a Δ follows from Lemma A.24.

For any $\Delta < \bar{\Delta} = \min\{\bar{\Delta}_1, \Delta_2\}$, let $O^{\Delta *}$ be among the optimal solutions. Define $\mathbf{N}^{\Delta} = \left\{N_i^{\Delta}\right\}_{i \in I^{K(\Delta)}}$ to be such that $N_i^{\Delta} = \left|O^{\Delta *} \cap \mathcal{S}_i^{K(\Delta)}\right|$. Similarly, define $\tilde{\mathbf{N}}^{\Delta} = \left\{\tilde{N}_i^{\Delta}\right\}_{i \in I^{K(\Delta)}}$ where $\tilde{N}_i^{\Delta} = \min\left\{N_i^{\Delta}, \left\lceil \frac{K(\Delta)^{\mathsf{d}}}{\Delta^{\mathsf{d}}}\bar{n}\right\rceil\right\}$.

Toward a contradiction, suppose that $\Delta^{\mathsf{d}}\left|O^{\Delta *}\right| \geq \bar{N} + m$. The first step of the proof implies

$$\bar{U}^{K(\Delta),\Delta}\left(\mathbf{N}^{\Delta}\right) - \bar{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right) \leq -\frac{R}{3}m$$

Finally, we have that $\Pi^{\Delta}\left(O^{\Delta *}\right) = U^{K(\Delta),\Delta}\left(\mathbf{N}^{\Delta}\right) \leq \bar{U}^{K(\Delta),\Delta}\left(\mathbf{N}^{\Delta}\right)$, and $U^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right) \geq \underline{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right)$. Together, these imply that if $\Delta < \bar{\Delta}$,

$$\begin{aligned}
U^{K(\Delta),\Delta}\left(\mathbf{N}^{\Delta}\right) - U^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right) &\leq \bar{U}^{K(\Delta),\Delta}\left(\mathbf{N}^{\Delta}\right) - \underline{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right) \\
&= \bar{U}^{K(\Delta),\Delta}\left(\mathbf{N}^{\Delta}\right) - \bar{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right) + \bar{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right) - \underline{U}^{K(\Delta),\Delta}\left(\tilde{\mathbf{N}}^{\Delta}\right) \\
&\leq -\frac{R}{3}m + \frac{R}{4}m \\
&< 0
\end{aligned}$$

Therefore, $O^{\Delta *}$ cannot be optimal, a contradiction. ■

Lemma A.26 *Fix $\epsilon > 0$. There exists a $\bar{\Delta}$ such that for any $\Delta < \bar{\Delta}$ and any $\mathbf{N} = \{N_i\}_{i \in I^{K(\Delta)}}$ such that $\sum_{i \in I^{K(\Delta)}} N_i \leq 2\bar{N}$,*

$$\left|U^{K(\Delta),\Delta}(\mathbf{N}) - \Pi\left(\hat{n}^{K(\Delta),\Delta}(\mathbf{N})\right)\right| < \epsilon$$

Proof. The first step in this proof is to show that for any k, Δ, and any admissible $\mathbf{N} = \{N_i\}_{i \in I^k}$, we can bound $\left|U^{k,\Delta}(\mathbf{N}) - \Pi\left(\hat{n}^{k,\Delta}(\mathbf{N})\right)\right|$. We already know that

$$\underline{U}^{k,\Delta}(\mathbf{N}) \leq U^{k,\Delta}(\mathbf{N}) \leq \bar{U}^{k,\Delta}(\mathbf{N})$$

In addition, we have

$$\underline{U}^{k,\Delta}(\mathbf{N}) \leq \Pi\left(\hat{n}^{k,\Delta}(\mathbf{N})\right) \leq \bar{U}^{k,\Delta}(\mathbf{N})$$

because

$$\begin{aligned}
\Pi\left(\hat{n}^{k,\Delta}(\mathbf{N})\right) &= \max_n \int_{s\in\mathcal{S}} \left\{-R_s n_s + D_s b_s z\left(q, \sum_{i\in I^k} N_i\right)^{\varepsilon-1} \kappa(n_s)\right\} ds \;\; \text{subject to} \int_{s\in\mathcal{S}_i^k} n_s ds = N_i \\
&\leq \max_n \int_{s\in\mathcal{S}} \left\{-\underline{R}_i^k n_s + \bar{D}_i^k \bar{b}_i^k z\left(q, \sum_{i\in I^k} N_i\right)^{\varepsilon-1} \kappa(n_s)\right\} ds \;\; \text{subject to} \int_{s\in\mathcal{S}_i^k} n_s ds = N_i \\
&\leq \bar{U}^{k,\Delta}(\mathbf{N})
\end{aligned}$$

and

$$\begin{aligned}
\Pi\left(\hat{n}^{k,\Delta}(\mathbf{N})\right) &= \max_n \int_{s\in\mathcal{S}} \left\{-R_s n_s + D_s b_s z\left(q, \sum_{i\in I^k} N_i\right)^{\varepsilon-1} \kappa(n_s)\right\} ds \;\; \text{subject to} \int_{s\in\mathcal{S}_i^k} n_s ds = N_i \\
&\geq \max_n \int_{s\in\mathcal{S}} \left\{-\bar{R}_i^k n_s + \underline{D}_i^k \underline{b}_i^k z\left(q, \sum_{i\in I^k} N_i\right)^{\varepsilon-1} \kappa(n_s)\right\} ds \;\; \text{subject to} \int_{s\in\mathcal{S}_i^k} n_s ds = N_i \\
&\geq \underline{U}^{k,\Delta}(\mathbf{N})
\end{aligned}$$

Together, these imply that

$$\left|U^{K(\Delta),\Delta}(\mathbf{N}) - \Pi\left(\hat{n}^{K(\Delta),\Delta}(\mathbf{N})\right)\right| \leq \left|\bar{U}^{K(\Delta),\Delta}(\mathbf{N}) - \underline{U}^{K(\Delta),\Delta}(\mathbf{N})\right|$$

Finally, Lemma A.24 gives that there is a $\bar{\Delta}$ such that $\left|\bar{U}^{K(\Delta),\Delta}(\mathbf{N}) - \underline{U}^{K(\Delta),\Delta}(\mathbf{N})\right| \leq \epsilon$, and hence

$$\left|U^{K(\Delta),\Delta}(\mathbf{N}) - \Pi\left(\hat{n}^{K(\Delta),\Delta}(\mathbf{N})\right)\right| < \epsilon$$

■

Lemma A.27 *For any $\epsilon > 0$, there is a $\bar{\Delta}$ small enough so that $\Delta < \bar{\Delta}$ implies*

$$\sum_{i\in I^{K(\Delta)}} \left|\Delta^{\mathtt{d}}\left|O^{\Delta*}\cap\mathcal{S}_i^{K(\Delta)}\right| - \int_{s\in\mathcal{S}_i^{K(\Delta)}} n_s^* ds\right| < \epsilon$$

Proof. For any Δ, define $n^{\Delta*} \equiv \arg\max_n \Pi(n)$ subject to $\int_{s\in\mathcal{S}_i^{K(\Delta)}} n_s ds = \Delta^{\mathtt{d}}\left|O^{\Delta*}\cap\mathcal{S}_i^{K(\Delta)}\right|$.

We first show that, for any $\eta > 0$, there is an $\bar{\Delta}$ small enough so that $\Delta < \bar{\Delta}$ implies that $\left|\Pi(n^*) - \Pi\left(n^{\Delta*}\right)\right| \leq$

η. For any Δ, optimality implies both

$$
\begin{aligned}
0 &\leq \Pi(n^*) - \Pi\left(n^{\Delta*}\right) \\
0 &\leq U^{K(\Delta),\Delta}\left(\left\{\left|O^{\Delta*} \cap \mathcal{S}_i^{K(\Delta)}\right|\right\}_{i \in I^{K(\Delta)}}\right) - U^{K(\Delta),\Delta}\left(\left\{\frac{1}{\Delta^{\mathsf{d}}} \int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^* ds\right\}_{i \in I^{K(\Delta)}}\right)
\end{aligned}
$$

Adding these together and rearranging gives

$$
\begin{aligned}
0 &\leq \Pi(n^*) - \Pi\left(n^{\Delta*}\right) + U^{K(\Delta),\Delta}\left(\left\{\left|O^{\Delta*} \cap \mathcal{S}_i^{K(\Delta)}\right|\right\}_{i \in I^{K(\Delta)}}\right) - U^{K(\Delta),\Delta}\left(\left\{\frac{1}{\Delta^{\mathsf{d}}} \int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^* ds\right\}_{i \in I^{K(\Delta)}}\right) \\
&= \Pi(n^*) - U^{K(\Delta),\Delta}\left(\left\{\frac{1}{\Delta^{\mathsf{d}}} \int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^* ds\right\}_{i \in I^{K(\Delta)}}\right) - \Pi\left(n^{\Delta*}\right) + U^{K(\Delta),\Delta}\left(\left\{\left|O^{\Delta*} \cap \mathcal{S}_i^{K(\Delta)}\right|\right\}_{i \in I^{K(\Delta)}}\right) \\
&\leq \left|\Pi(n^*) - U^{K(\Delta),\Delta}\left(\left\{\frac{1}{\Delta^{\mathsf{d}}} \int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^* ds\right\}_{i \in I^{K(\Delta)}}\right)\right| + \left|\Pi\left(n^{\Delta*}\right) - U^{K(\Delta),\Delta}\left(\left\{\left|O^{\Delta*} \cap \mathcal{S}_i^{K(\Delta)}\right|\right\}_{i \in I^{K(\Delta)}}\right)\right|
\end{aligned}
$$

Note that $\int_{s \in \mathcal{S}} n_s^* ds \leq \bar{N}$ and, for sufficiently small Δ, $\Delta^{\mathsf{d}}\left|O^{\Delta*}\right| \leq 2\bar{N}$ (the latter uses Lemma A.25 with $m = \bar{N}$). Lemma A.26 therefore implies that there is a $\bar{\Delta}$ such that $\Delta < \bar{\Delta}$ implies that each term is less than $\frac{\eta}{2}$. As a result, $0 \leq \Pi(n^*) - \Pi\left(n^{\Delta*}\right) \leq \eta$.

Claim A.23 states that for any $\epsilon > 0$, there is an η such that $\left|\Pi(n^*) - \Pi\left(n^{\Delta*}\right)\right| < \eta$ implies $\left\|n^* - n^{\Delta*}\right\| < \epsilon$. Thus there is an $\bar{\Delta}$ small enough so that $\Delta < \bar{\Delta}$ implies that $\left|\Pi(n^*) - \Pi\left(n^{\Delta*}\right)\right| < \eta$ and hence $\left\|n^* - n^{\Delta*}\right\| < \epsilon$. Finally, the definition of $n^{\Delta*}$ implies $\int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^{\Delta*} ds = \Delta^{\mathsf{d}}\left|O^{\Delta*} \cap \mathcal{S}_i^{K(\Delta)}\right|$. Thus if $\Delta < \bar{\Delta}$, so we have

$$
\begin{aligned}
\sum_{i \in I^{K(\Delta)}} \left|\Delta^{\mathsf{d}}\left|O^{\Delta*} \cap \mathcal{S}_i^{K(\Delta)}\right| - \int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^* ds\right| &= \sum_{i \in I^{K(\Delta)}} \left|\int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^{\Delta*} ds - \int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^* ds\right| \\
&\leq \sum_{i \in I^{K(\Delta)}} \int_{s \in \mathcal{S}_i^{K(\Delta)}} \left|n_s^{\Delta*} - n_s^*\right| ds \\
&= \left\|n^{\Delta*} - n^*\right\| \\
&< \epsilon
\end{aligned}
$$

■

Proposition A.28 *Consider any Jordan measurable set $\mathcal{A}$. For any ϵ, there is a $\bar{\Delta}$ such that $\Delta < \bar{\Delta}$ implies that $\left|\Delta^{\mathsf{d}}\left|O^{\Delta*} \cap \mathcal{A}\right| - \int_{s \in \mathcal{A}} n_s^* ds\right| < \epsilon$.*

Proof. Let $\bar{I}^k(\mathcal{A}) = \left\{i \in I^k \text{ such that } \mathcal{S}_i^k \cap \mathcal{A} \neq \emptyset\right\}$ be the smallest collection of squares/segments that contains $\mathcal{A}$, and let $\underline{I}^k(\mathcal{A}) = \left\{i \in I^k \text{ such that } \mathcal{S}_i^k \subseteq \mathcal{A}\right\}$ be the largest collection contained in A. Let $\mathcal{B}^k(\mathcal{A}) = \cup_{i \in \bar{I}^k(\mathcal{A}) \setminus \underline{I}^k(\mathcal{A})} \mathcal{S}_i^k$ be the union of segments/squares that contains points in both $\mathcal{A}$ and its complement.

Consider and optimal solution $O^{\Delta *}$ for economy Δ. For any k, define $N_i^{k\Delta} \equiv \left|O^{\Delta *} \cap S_i^k\right|$ to be the number of plants in segment/square S_i^k. We have

$$\sum_{i \in \underline{I}^k(\mathcal{A})} N_i^{k\Delta} \leq \left|O^{\Delta *} \cap \mathcal{A}\right| \leq \sum_{i \in \bar{I}^k(\mathcal{A})} N_i^{k\Delta}$$

By multiplying through by $\Delta^{\mathtt{d}}$ and subtracting $\int_{s \in \mathcal{A}} n_s^* ds$ from each side, the left hand inequality can be expressed as

$$\begin{aligned}
\Delta^{\mathtt{d}}\left|O^{\Delta *} \cap \mathcal{A}\right| - \int_{s \in \mathcal{A}} n_s^* ds &\geq \sum_{i \in \underline{I}^k(\mathcal{A})} \Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{A}} n_s^* ds \\
&\geq \sum_{i \in \underline{I}^k(\mathcal{A})} \left(\Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{S}_i^k} n_s^* ds\right) - \int_{s \in \mathcal{B}^k(\mathcal{A})} n_s^* ds \\
&\geq -\sum_{i \in \underline{I}^k(\mathcal{A})} \left|\Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{S}_i^k} n_s^* ds\right| - \int_{s \in \mathcal{B}^k(\mathcal{A})} n_s^* ds \\
&\geq -\sum_{i \in \bar{I}^k(\mathcal{A})} \left|\Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{S}_i^k} n_s^* ds\right| - \int_{s \in \mathcal{B}^k(\mathcal{A})} n_s^* ds
\end{aligned}$$

Similarly, by multiplying through by $\Delta^{\mathtt{d}}$ and subtracting $\int_{s \in \mathcal{A}} n_s^* ds$ from each side, the right hand inequality can be expressed as

$$\begin{aligned}
\Delta^{\mathtt{d}}\left|O^{\Delta *} \cap \mathcal{A}\right| - \int_{s \in \mathcal{A}} n_s^* ds &\leq \sum_{i \in \bar{I}^k(\mathcal{A})} \Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{A}} n_s^* ds \\
&\leq \sum_{i \in \bar{I}^k(\mathcal{A})} \left(\Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{S}_i^k} n_s^* ds\right) + \int_{s \in \mathcal{B}^k(\mathcal{A})} n_s^* ds \\
&\leq \sum_{i \in \bar{I}^k(\mathcal{A})} \left|\Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{S}_i^k} n_s^* ds\right| + \int_{s \in \mathcal{B}^k(\mathcal{A})} n_s^* ds
\end{aligned}$$

Together, these give

$$\begin{aligned}
\left|\Delta^{\mathtt{d}}\left|O^{\Delta *} \cap \mathcal{A}\right| - \int_{s \in \mathcal{A}} n_s^* ds\right| &\leq \sum_{i \in \bar{I}^k(\mathcal{A})} \left|\Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{S}_i^k} n_s^* ds\right| + \int_{s \in \mathcal{B}^k(\mathcal{A})} n_s^* ds \\
&\leq \sum_{i \in I^k} \left|\Delta^{\mathtt{d}} N_i^{k\Delta} - \int_{s \in \mathcal{S}_i^k} n_s^* ds\right| + \bar{n}\left|\mathcal{B}^k(\mathcal{A})\right|
\end{aligned}$$

In particular this bound holds for any Δ, $K(\Delta)$ pair.

From Lemma A.27, there is a $\bar{\Delta}_1$ small enough so that $\Delta < \bar{\Delta}_1$ implies that

$$\sum_{i \in I^{K(\Delta)}} \left| \Delta^{\mathsf{d}} N_i^{K(\Delta)\Delta} - \int_{s \in \mathcal{S}_i^{K(\Delta)}} n_s^* ds \right| < \frac{\epsilon}{2}.$$

Further, since $\mathcal{A}$ is Jordan measurable, $\lim_{k \to 0} \mathcal{B}^k(\mathcal{A}) = 0$, so there is a $\bar{\Delta}_2$ small enough so that $\Delta < \bar{\Delta}_2$ implies $\left|\mathcal{B}^{K(\Delta)}(\mathcal{A})\right| < \frac{1}{\bar{n}}\frac{\epsilon}{2}$. Together, these imply that $\Delta < \min\left\{\bar{\Delta}_1, \bar{\Delta}_2\right\}$ implies that

$$\left| \Delta^{\mathsf{d}} \left| O^{\Delta *} \cap \mathcal{A} \right| - \int_{s \in \mathcal{A}} n_s^* ds \right| < \epsilon.$$

■

A.4 Additional Proofs

We formally state and prove here the additional result, quoted in the text, about the relative marginal efficiency of distribution.

Lemma A.29 *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

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

■

A.4.1 Proof of Proposition 6

Proof. $n_s(z)$ denotes the density of plants a firm with productivity z places in location s. The first order condition (4) implies that $n_s(z) = 0$ if $\frac{R_s+\lambda_j}{x_s z^{\varepsilon-1}} > \kappa'(0)$. Since $\lambda_j > 0$,

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

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

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

The result follows from the fact that κ' is continuous, strictly decreasing, and $\kappa'(0) < \infty$ if $\lim_{\delta \to \infty} \frac{\delta^{\mathsf{d}}}{t(\delta)^{\varepsilon-1}} = 0$ by Lemma 3. ■

A.4.2 Proof of Lemma 7

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 $L_j = \frac{\varepsilon-1}{W}\int x_s z_j^{\varepsilon-1}\kappa(n_{js})\,ds = \frac{\sigma}{W}\lambda_j$. If $N_1 = N_2 = 0$, then $L_1 = L_2 = 0$. Otherwise, by Lemma 4, $\frac{\lambda_2}{z_2^{\varepsilon-1}} > \frac{\lambda_1}{z_1^{\varepsilon-1}}$, which implies $\lambda_2 > \lambda_1$, and so $L_2 > L_1$. ■

A.5 Proof of Proposition 8: Aggregation

In this appendix we prove proposition Proposition 8 and show some additional aggregate properties of the industry equilibrium defined in Section 3.

A.5.1 The Local Price Index

The price that firm j sets in location s is

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

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).[43]

[43] 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.

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 $Z\left(q,N\right) \equiv z\left(q,\Delta^{\mathtt{d}}N\right)^{\varepsilon-1}$ gives

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

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

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

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.5.2 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

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

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

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

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.5.3 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

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

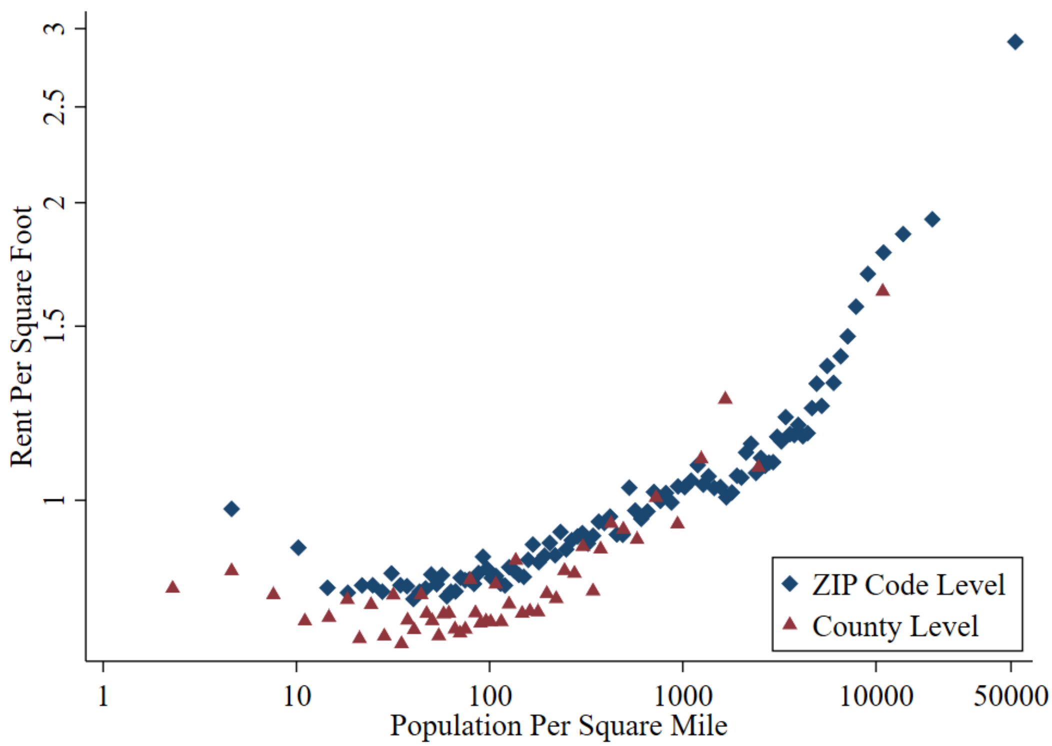
Combining this with the expression for the price level $P_s=\frac{\varepsilon}{\varepsilon-1}\frac{W_s}{B_s\mathcal{Z}_s}$ and simplifying gives

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

B Rents and Density

Our theory relates the sorting of firms to local rent. Data on rent is incomplete and available at irregular geographic units, so our empirical exercises relate sorting to population density. Nevertheless, we can test empirically 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 (Manson et al. (2021)). 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) from Manson et al. (2021) to construct the population density measure.

C Sorting

In this section we perform several robustness checks to our sorting results.

Table V presents the results of regressing the average of 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, on the log of the national size of the firm and industry fixed effects. The first panel is the analogous to Figure 8 and presents the results as we vary M. Panels 2 to 7 use $M = 12$. The second panel restricts the analysis to firms with at least X plants. The third panel adds to the second panel a headquarters' location fixed effect for each firm. The fourth panel restricts the analysis to industries where there is a firm with at least X plants, and the fifth panel adds the fixed effect for the headquarters' location.

The sixth panel repeats the analysis by major industry. The last panel shows the robustness of the baseline results to excluding the own firm contribution to employment density, alternative weighting schemes, and to using only non-imputed data.

Table V: Sorting: Firm Size and Local Density

	(1) $\ln \bar{L}_j$	(2) $\ln \bar{L}_j$	(3) $\ln \bar{L}_j$	(4) $\ln \bar{L}_j$	(5) $\ln \bar{L}_j$
Baseline					
$\ln L_j$	0.222***	0.196***	0.165***	0.129***	0.0952***
	(0.00104)	(0.00101)	(0.000975)	(0.000926)	(0.000848)
Observations	3,645,763	3,665,497	3,670,994	3,672,721	3,673,053
R-squared	0.159	0.153	0.139	0.120	0.099
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	3	6	12	24	48
Firms with at least X plants	*Baseline*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln L_j$	0.165***	0.134***	0.132***	0.135***	0.146***
	(0.000975)	(0.00655)	(0.00982)	(0.0153)	(0.0249)
Observations	3,670,994	11,203	4,904	1,892	876
R-squared	0.139	0.385	0.391	0.405	0.384
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
Controlling for HQ location, firms with at least X plants	*X = 2*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln L_j$	0.0229***	0.0347***	0.0396***	0.0516***	0.0791**
	(0.00171)	(0.00563)	(0.00863)	(0.0159)	(0.0350)
Observations	145,186	9,700	4,182	1,534	652
R-squared	0.705	0.665	0.676	0.693	0.664
SIC8 FE	Yes	Yes	Yes	Yes	Yes
HQ Location FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
Industries where largest firm has at least X plants	*Baseline*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln L_j$	0.165***	0.178***	0.176***	0.176***	0.172***
	(0.000975)	(0.00109)	(0.00119)	(0.00140)	(0.00164)
Observations	3,670,994	2,861,609	2,424,907	1,829,818	1,387,742
R-squared	0.139	0.114	0.102	0.091	0.080
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
Controlling for HQ location, industries where largest firm has at least X plants	*Baseline*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln L_j$	0.0229***	0.0252***	0.0244***	0.0222***	0.0201***
	(0.00171)	(0.00179)	(0.00190)	(0.00214)	(0.00244)
Observations	145,186	124,065	106,163	81,372	62,957
R-squared	0.705	0.711	0.720	0.729	0.735
SIC8 FE	Yes	Yes	Yes	Yes	Yes
HQ Location FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
	Baseline	*Excluding Own Contribution*	*Unweighted*	*Weighted by Employment*	*Non-imputed*
$\ln L_j$	0.165***	0.164***	0.161***	0.164***	0.202***
	(0.000975)	(0.000975)	(0.000974)	(0.000981)	(0.00111)
Observations	3,670,994	3,668,734	3,670,994	3,670,994	2,605,050
R-squared	0.139	0.138	0.138	0.139	0.169
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12

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

Notes: The table presents the results of regressing the log of the average employment density across all of the firm plants on the log employment of the firm at the national level and industry fixed effects. The first panel presents the results as we vary M. Panels 2 to 7 use $M = 12$. The second panel restricts the analysis to firms with at least X plants. The third panel adds to the second panel a headquarters' location fixed effect for each firm. The fourth panel restricts the analysis to industries where there is a firm with at least X plants, and the fifth panel adds the fixed effect for the headquarters' location. The sixth panel repeats the analysis by major industry. The last panel shows the robustness of the baseline results to excluding the own firm contribution to employment density, alternative weighting schemes, and to using only non-imputed data.

D The Largest Firm in Town

Table VI presents the results of regressing the log of the national size of the firm with most plants in a location, $L_{j^*(s)}$, on the log of the employment density of the location, $\mathcal{L}_s$, and industry fixed effects. If there is a tie in the identity of the firm with most plants in a location, we take the average of the log national employment of the firms. The first panel presents our baseline results for different spatial resolutions M. The second panel repeats the analysis but restricting to firms with at least X plants, and the third panel restricts the analysis to industries where the largest firm has at least X plants. The fourth panel presents the results by major industry. The fifth panel presents the results when excluding the firm's own contribution to a location employment, when using alternative ways to resolving ties (in terms of which firm has the highest amount of plants in a location), and when using only non-imputed data.

Table VI: The national size of the largest firm in town

	(1)	(2)	(3)	(4)	(5)
	$\ln L_{j^*(s)}$	$\ln L_{j^*(s)}$	$\ln L_{j^*(s)}$	$\ln L_{j^*(s)}$	$\ln L_{j^*(s)}$
Baseline					
$\ln \mathcal{L}_s$	0.194***	0.293***	0.395***	0.486***	0.594***
	(0.00169)	(0.00206)	(0.00266)	(0.00350)	(0.00461)
Observations	3,131,324	2,551,226	1,984,474	1,473,278	1,006,305
R-squared	0.593	0.608	0.616	0.630	0.644
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	3	6	12	24	48
Firms with at least X plants	*Baseline*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln \mathcal{L}_s$	0.395***	0.239***	0.191***	0.158***	0.131***
	(0.00266)	(0.00620)	(0.00659)	(0.00729)	(0.00811)
Observations	1,984,474	356,238	308,839	253,996	211,517
R-squared	0.616	0.561	0.564	0.560	0.554
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
Industries where largest firm has at least X plants	*Baseline*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln \mathcal{L}_s$	0.395***	0.411***	0.430***	0.471***	0.516***
	(0.00266)	(0.00276)	(0.00291)	(0.00325)	(0.00366)
Observations	1,984,474	1,390,883	1,125,690	813,539	616,248
R-squared	0.616	0.613	0.609	0.605	0.600
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
By major industry	*Baseline*	*Manufacturing*	*Services*	*Retail Trade*	*FIRE*
$\ln \mathcal{L}_s$	0.395***	0.0746***	0.279***	0.562***	0.557***
	(0.00266)	(0.00549)	(0.00489)	(0.00491)	(0.00729)
Observations	1,984,474	245,343	647,569	421,352	133,956
R-squared	0.616	0.256	0.347	0.679	0.557
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
	Baseline	*Excluding Own Contribution*	*Discarding Ties*	*Largest Firm Among Ties*	*Non-imputed*
$\ln \mathcal{L}_s$	0.395***	0.372***	0.617***	0.694***	0.404***
	(0.00266)	(0.00491)	(0.00386)	(0.00276)	(0.00288)
Observations	1,984,474	568,124	1,449,578	1,984,474	1,666,909
R-squared	0.616	0.550	0.604	0.585	0.618
SIC8 FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12

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

Notes: The table presents the results of regressing the log of the national size of the firm with most plants in a location on the log of the employment density of the location and industry fixed effects. The first panel presents the results as we vary M. Panels 2 to 5 use $M = 12$. The second panel presents the results for firms with at least X plants. The third panel presents the results for industries with at least one firm with X plants. The fourth panel presents the results by major industry. The fifth panel presents the results when excluding the firm's own contribution to employment in a location, when discarding locations where there is a tie in the identity of the firm with most plants or using the size of the largest firm in this case, and when using only non-imputed data.

E Span of Control

In this section we present several robustness checks to our span-of-control results. Table VII presents the results of regressing the log of the average plant employment of a firm within a location on the firm's log national employment, and controlling for the firm's log number of plants in the location and square of the log number of plants in the location. In all cases, we subtract the own firm contribution of employment in a location from that firm's total employment. The first panel presents the results for different values of M. The second panel restricts the analysis to firms with at least X plants, while the third panel restricts the analysis to industries where there is one firm with at least X plants. The fourth panel presents the results by major industry. Table VIII presents some additional robustness results. The first panel presents the regression results without subtracting the own firm contribution of employment in a location from that firm's total employment. Panels 2 and 3 subtract the own firm contribution of employment in a location from that firm's total employment. Panel 2 adds higher order terms of the log of the number of establishments of a firm in a location as controls, while Panel 3 restricts attention to non-imputed data.

Table VII: Span of Control

	(1) $\ln \bar{l}_{js}$	(2) $\ln \bar{l}_{js}$	(3) $\ln \bar{l}_{js}$	(4) $\ln \bar{l}_{js}$	(5) $\ln \bar{l}_{js}$
Baseline					
$\ln L_{j,-js}$	0.0925***	0.103***	0.114***	0.123***	0.131***
	(0.00101)	(0.000918)	(0.000897)	(0.000911)	(0.000962)
$\ln n_{js}$	0.131***	0.109***	0.137***	0.163***	0.172***
	(0.0171)	(0.0114)	(0.00897)	(0.00753)	(0.00682)
$(\ln n_{js})^2$	0.00245	-0.0502***	-0.0813***	-0.0852***	-0.0811***
	(0.0135)	(0.00715)	(0.00447)	(0.00316)	(0.00251)
Observations	311,244	376,723	409,364	408,521	386,094
R-squared	0.632	0.608	0.573	0.542	0.511
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
M	3	6	12	24	48
Firms with at least X plants	*Baseline*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln L_{j,-js}$	0.114***	0.154***	0.187***	0.233***	0.275***
	(0.000897)	(0.00157)	(0.00190)	(0.00241)	(0.00296)
$\ln n_{js}$	0.137***	-0.0121	-0.0765***	-0.132***	-0.168***
	(0.00897)	(0.00960)	(0.00990)	(0.0104)	(0.0111)
$(\ln n_{js})^2$	-0.0813***	-0.0339***	-0.0188***	-0.00932*	-0.00158
	(0.00447)	(0.00457)	(0.00465)	(0.00494)	(0.00528)
Observations	409,364	233,744	197,273	157,980	126,999
R-squared	0.573	0.658	0.686	0.714	0.746
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
Industries where largest firm has at least X plants	*Baseline*	*X = 10*	*X = 20*	*X = 50*	*X = 100*
$\ln L_{j,-js}$	0.114***	0.114***	0.112***	0.108***	0.104***
	(0.000897)	(0.000897)	(0.000898)	(0.000901)	(0.000911)
$\ln n_{js}$	0.137***	0.133***	0.127***	0.102***	0.0720***
	(0.00897)	(0.00895)	(0.00894)	(0.00888)	(0.00883)
$(\ln n_{js})^2$	-0.0813***	-0.0797***	-0.0768***	-0.0673***	-0.0564***
	(0.00447)	(0.00446)	(0.00444)	(0.00438)	(0.00431)
Observations	409,364	405,623	394,807	369,321	336,424
R-squared	0.573	0.573	0.574	0.577	0.588
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12
By major industry	*Baseline*	*Manufacturing*	*Services*	*Retail Trade*	*FIRE*
$\ln L_{j,-js}$	0.114***	0.154***	0.130***	0.124***	0.0627***
	(0.000897)	(0.0104)	(0.00214)	(0.00118)	(0.00180)
$\ln n_{js}$	0.137***	1.127***	0.325***	-0.0506***	0.126***
	(0.00897)	(0.145)	(0.0281)	(0.0109)	(0.0159)
$(\ln n_{js})^2$	-0.0813***	-0.252**	-0.121***	-0.0451***	-0.0398***
	(0.00447)	(0.102)	(0.0164)	(0.00544)	(0.00720)
Observations	409,364	8,864	95,301	164,941	84,798
R-squared	0.573	0.541	0.506	0.665	0.442
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
M	12	12	12	12	12

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

Notes: The table presents the results of regressing the log of the average plant employment of a firm within a location on the log national employment of the firm (excluding the own firm contribution of employment in a location from that firm's total employment), industry fixed effects and controls for the number of plants that the firm has in the location. The first panel presents the results for different values of M. Panels 2 to 4 use $M = 12$. The second panel restricts the analysis to firms with at least X plants, while the third panel restricts the analysis to industries where there is one firm with at least X plants. The fourth panel presents the results by major industry. In all cases, we subtract the own firm contribution of employment in a location from that firm's total employment.

Table VIII: Span of Control: Additional Exercises

	(1)	(2)	(3)	(4)	(5)
	$\ln \bar{l}_{js}$	$\ln \bar{l}_{js}$	$\ln \bar{l}_{js}$	$\ln \bar{l}_{js}$	$\ln \bar{l}_{js}$
$\ln L_j$	0.216***	0.240***	0.277***	0.318***	0.371***
	(0.000531)	(0.000495)	(0.000501)	(0.000528)	(0.000571)
$\ln n_{js}$	-0.367***	-0.548***	-0.636***	-0.714***	-0.814***
	(0.0156)	(0.00924)	(0.00679)	(0.00558)	(0.00486)
$(\ln n_{js})^2$	0.0807***	0.0737***	0.0544***	0.0560***	0.0605***
	(0.0131)	(0.00598)	(0.00341)	(0.00240)	(0.00187)
Observations	2,033,197	2,574,018	3,009,650	3,343,237	3,592,041
R-squared	0.597	0.568	0.550	0.544	0.552
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
M	3	6	12	24	48
$\ln L_{j,-js}$	0.0925***	0.103***	0.114***	0.123***	0.131***
	(0.00101)	(0.000918)	(0.000897)	(0.000911)	(0.000963)
$\ln n_{js}$	-0.0310	0.206***	0.319***	0.319***	0.339***
	(0.0497)	(0.0335)	(0.0256)	(0.0211)	(0.0186)
$(\ln n_{js})^2$	0.268***	-0.211***	-0.344***	-0.278***	-0.274***
	(0.0890)	(0.0570)	(0.0398)	(0.0297)	(0.0235)
$(\ln n_{js})^3$	-0.102**	0.0688**	0.0953***	0.0575***	0.0559***
	(0.0430)	(0.0278)	(0.0183)	(0.0125)	(0.00901)
$(\ln n_{js})^4$	0.00909	-0.00757**	-0.00840***	-0.00367**	-0.00385***
	(0.00558)	(0.00385)	(0.00248)	(0.00156)	(0.00103)
Observations	311,244	376,723	409,364	408,521	386,094
R-squared	0.632	0.608	0.573	0.542	0.512
SIC8-location FE	Yes	Yes	Yes	Yes	Yes
M	3	6	12	24	48
Non-Imputed					
$\ln L_{j,-js}$	0.0854***	0.0947***	0.106***	0.115***	0.125***
	(0.00114)	(0.00104)	(0.00101)	(0.00102)	(0.00107)
$\ln n_{js}$	0.193***	0.153***	0.154***	0.174***	0.173***
	(0.0208)	(0.0138)	(0.0104)	(0.00863)	(0.00770)
$(\ln n_{js})^2$	-0.0165	-0.0546***	-0.0777***	-0.0830***	-0.0795***
	(0.0169)	(0.00898)	(0.00527)	(0.00372)	(0.00294)
Observations	227,058	280,597	310,989	317,407	306,306
R-squared	0.647	0.625	0.592	0.563	0.535
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 log of the average plant employment of a firm within a location on the log national employment of the firm, industry fixed effects and controls for the number of plants that the firm has in the location. The first panel presents the results for different values of M, without excluding the own firm contribution of employment in a location from that firm's total employment. Panels 2 and 3 exclude the own firm contribution of employment in a location from that firm's total employment. Panel 2 includes higher order terms of the log of the number of establishments of a firm in a location as control variables. Panel 3 restricts attention to non-imputed data.

F 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 4 and 5),

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

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 4), (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 $S \times J$ 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 3). 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.[44] 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 $\mathcal{H}(0) = 0$, $\lim_{n\to\infty}\mathcal{H}(n) = 1$.[45] 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 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,

$$\begin{aligned}
\text{absolute error } &= \int_s \int_j \mathbf{1}\left[\hat{n}_{js} > 0\right] \overbrace{\left[\hat{x}_s \hat{z}_j^{\varepsilon-1}\kappa'(\hat{n}_{js}) - (R_s + \hat{\lambda}_j)\right]}^{\text{error in equation 4}} dj ds \ , \\
\text{relative error } &= \frac{\int_s \int_j \mathbf{1}\left[\hat{n}_{js} > 0\right] \left[\hat{x}_s \hat{z}_j^{\varepsilon-1}\kappa'(\hat{n}_{js}) - (R_s + \hat{\lambda}_j)\right] dj ds}{\int_s \int_j \mathbf{1}\left[\hat{n}_{js} > 0\right](R_s + \hat{\lambda}_j) dj ds} \ .
\end{aligned}$$

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 4. 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.

G Numerical Computation of Bounds $\bar{\pi}_j^{k\Delta}$ and $\underline{\pi}_j^{k\Delta}$

In this section we provide a numerical example of the upper and lower bounds $\bar{\pi}_j^{k\Delta}$ and $\underline{\pi}_j^{k\Delta}$ for any given k and Δ. We then use this example to discuss how the gap between these bounds changes with Δ.

We begin by noticing that the expression for the upper bound $\bar{\pi}_j^{k\Delta}$ presented in Claim A.11 can be

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

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

written as

$$\bar{\pi}_j^{k\Delta} \equiv \sup_{\{n_i \geq 0\}} \sum_{i \in I^k} -n_i k^2 \underline{R}_i^k \xi + z \left(q_j, \sum_{i \in I^k} n_i k^2 \right)^{\varepsilon - 1} \sum_{i \in I^k} \bar{D}_i^k \bar{b}_i^k k^2 \kappa(n_i) + z\left(q_j, 0\right)^{\varepsilon-1} \bar{D}\bar{b} k^{-1} \int_0^k t\left(\frac{\delta}{\Delta}\right)^{1-\varepsilon} d\delta \ .$$

A similar expression can be obtained for $\underline{\pi}_j^{k\Delta}$. We can obtain expressions for n_i, the number of plants in square $\mathcal{S}_i^k$, from the first order condition,

$$z \left(q_j, \sum_{i \in I^k} n_i k^2 \right)^{\varepsilon - 1} \bar{D}_i^k \bar{b}_i^k \kappa'(n_i) = \underline{R}_i^k - \frac{d\left[z\left(q_j, \sum_{i \in I^k} n_i k^2\right)^{\varepsilon - 1} \right]}{dn_i} \sum_{i \in I^k} \bar{D}_i^k \bar{b}_i^k \kappa(n_i) \ .$$

Again, a similar first-order condition can be produced for the number of establishments in each location implied by the lower bound.

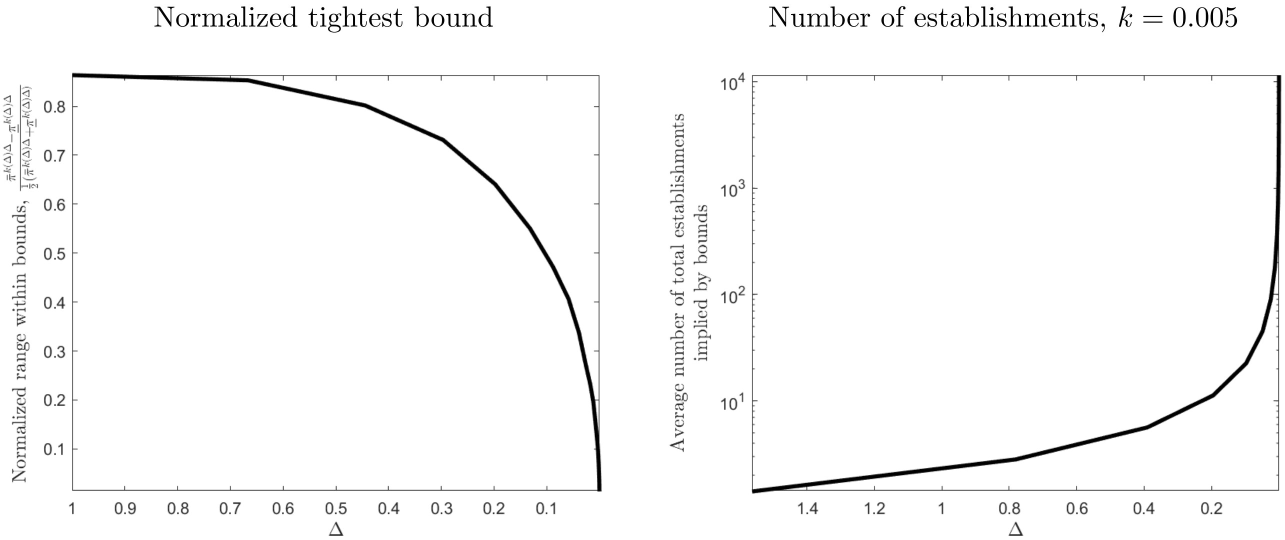
For given vectors $\{\bar{D}^k, \bar{b}^k, \underline{R}^k\}_{i \in I^k}$ and $\{\underline{D}^k, \underline{b}^k, R^k\}_{i \in I^k}$, we use the first order conditions to obtain $\bar{n}_i$ and $\underline{n}_i$ for firm j, the solutions to the two maximization problems. To do this, notice that the first order condition closely resembles that one presented in (4), and thus we exploit the procedure presented in Appendix F. With expressions for $\bar{n}_i$ and $\underline{n}_i$ we can then readily compute $\bar{\pi}_j^{k\Delta}$ and $\underline{\pi}_j^{k\Delta}$.

For the example presented below, we restrict our attention to space defined in a unit square, $\mathcal{S} = [0,1]^2$, and we assume that for any $\{s_1, s_2\} \in \mathcal{S}$ we have that $b_s = 50(1+s_1)(1+2s_2)$, $D_s = 20(1+s_1)(1+2s_2)$, and $R_s = 0.1(1+s_1)(1+s_2)$. Further, we set $\varepsilon = 2$, $q_j = 1$, and we follow Section 3.1 and assume that $t(\delta/\sqrt{\phi}) = e^{\delta/\sqrt{\phi}}$ with $\phi = 0.04$. Likewise, we set $z(q,N) = qe^{-N/\sigma}$ with $\sigma = 5$.

For a given k, we divide $\mathcal{S}$ into squares with side of length k. We do this for various values of k ranging from $1/2$ to $5/1000$. Because the number of partitions equals $(1/k)^2$, the number of squares in I^k ranges from 4 to 40,000. Then, for a fixed value of Δ, let $k(\Delta)$ denote which partition provides the tightest normalized range within the upper and lower bounds, $k(\Delta) = \arg\min_k \left(\bar{\pi}_j^{k\Delta} - \underline{\pi}_j^{k\Delta}\right) / \left[\frac{1}{2}\left(\bar{\pi}_j^{k\Delta} + \underline{\pi}_j^{k\Delta}\right)\right]$. The function $\left(\bar{\pi}_j^{k(\Delta)\Delta} - \underline{\pi}_j^{k(\Delta)\Delta}\right) / \left[\frac{1}{2}\left(\bar{\pi}_j^{k(\Delta)\Delta} + \underline{\pi}_j^{k(\Delta)\Delta}\right)\right]$ provides a good notion of the tightness of the bounds.

The left panel of Figure G.1 presents the normalized tightest bounds for the example we explore in this section, and the right panel presents the (log of the) average total number of establishments implied by the upper and lower bounds for the case $k = 0.005$. The figure shows that the bounds are not particularly tight when Δ is large, but they tighten as Δ falls. Consistent with our theoretical results, the bounds become very tight as Δ approaches zero. Likewise, the results presented in the right panel suggest that the bounds are tight when Δ is such that firms operate many plants, but less tight when Δ is large and the firm operates few plants. For example, if the firm were to operate over 1,000 establishments, the bounds appear to be very tight with a normalized gap close to zero. However, if the firm operates 10 total establishments, the normalized gap is around 0.7.

Figure G.1: Bounds



H A discrete example with firms with few plants

Our approach to the problem of choosing plant locations was to focus on the limiting economy as Δ approaches zero. In this limit, the problem admits an analytical solution that we use to derive theoretical predictions that we then corroborate empirically. While our theoretical results regarding uniform convergence of the policy function are reassuring of the relevance of these predictions for industries for which the limit is a good approximation, they may be less relevant for industries in which plants tend to have large catchment areas.

In this section, we use two numerical examples to explore firm choices outside of the limit, i.e. large Δ. In both examples, firms choose to have a small number of plants. In the first example, Δ is very large and firms choose to have either one or two plants, as transport costs are low. In this example, we run the same regressions as in the main text of the paper and fail to detect sorting. In the second example, we lower Δ and solve for the resulting plant configuration. Now firms place more plants across locations. In this example, we detect sorting that is consistent with the theoretical predictions of the limiting economy.

To make the numerical exercise feasible and operational we make a set of modifications to our model: (i) we follow Tintelnot (2016) and assume that a firm produces a continuum of goods, where each location that may be used to produce the good has a different idiosyncratic cost of producing a particular good, and (ii) we assume that there is only a discrete, and small, set of feasible locations. Modification (ii) allows us to use the toolkit in Arkolakis et al. (2017) to solve the plant location problem for each firm. Modification (i) increases the number of configurations that can be ruled out before resorting to evaluating all remaining combinations (the brute force approach).

Consider a set of discrete locations, where s denotes a location. Every firm produces a continuum of goods, $i \in [0,1]$. For each good, each location where the firm places a plant o has an idiosyncratic cost shifter a_{oi} with $\Pr[a_{oi} > a] = e^{-a^{\theta}}$, so that the unit cost of supplying good i from a plant in location o to location s is $\frac{w}{B_o Z} a_{oi} T_{os}$. Thus, the minimal cost to the firm of supplying good i to location s is $\lambda_{is} = \min_{o \in O} \frac{w}{B_o Z} a_{oi} T_{os}$, with associated distribution $\Pr(\lambda_{is} > \lambda) = e^{-\lambda^{\theta} \sum_{o \in O} \left(\frac{B_o Z}{w T_{os}}\right)^{\theta}}$.

Demand for each good at each location s is $D_s p^{-\varepsilon}$, which implies a markup of $\frac{\varepsilon}{\varepsilon - 1} > 1$. Given a set of active plants O, a firm's total profits are

$$\pi(O, Z) = \left(\frac{1}{\varepsilon - 1}\right) \left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \sum_s D_s \int_0^1 \lambda_{is}^{1-\varepsilon} di - \sum_{o \in O} R_o \xi \ ,$$

where, given the distribution for λ_{is}, $\int_0^1 \lambda_{is}^{1-\varepsilon} di = \Gamma\left(1 - \frac{\varepsilon - 1}{\theta}\right) \left[\sum_{o \in o} \left(\frac{B_o Z}{w T_{os}}\right)^{\theta}\right]^{\frac{\varepsilon-1}{\theta}}$. Therefore, a firm's profits from operating a set of plants O are given by

$$\pi(O, Z) = \left(\frac{1}{\varepsilon - 1}\right) \left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \Gamma\left(1 - \frac{\varepsilon - 1}{\theta}\right) \sum_s D_s \left[\sum_{o \in O} \left(\frac{B_o Z}{w T_{os}}\right)^{\theta}\right]^{\frac{\varepsilon-1}{\theta}} - \sum_{o \in O} R_o \xi \ ,$$

where plant employment in location o is given by

$$l_o = \frac{\varepsilon}{w} \left(\frac{\varepsilon}{\varepsilon - 1}\right)^{-\varepsilon} \Gamma\left(1 - \frac{\varepsilon - 1}{\theta}\right) \sum_s D_s \left[\sum_{\tilde{o} \in O} \left(\frac{B_{\tilde{o}} Z}{w T_{\tilde{o}s}}\right)^{\theta}\right]^{\frac{\varepsilon-1}{\theta}} \frac{\left(\frac{B_o Z}{w T_{os}}\right)^{\theta}}{\sum_{\tilde{o} \in O} \left(\frac{B_{\tilde{o}} Z}{w T_{\tilde{o}s}}\right)^{\theta}}.$$

Notice that we recover the specification in the main text, where each location in space is served by only one the firm's plants, when $\theta \to \infty$.[46] At each location s a firm with productivity q sells goods at price

$$p_s(q) = \frac{\varepsilon}{\varepsilon - 1} \left[\sum_{o \in O(q)} \left(\frac{B_o Z(q, N(q))}{w T_{os}}\right)^{\theta}\right]^{-\frac{1}{\theta}} .$$

As a result, the price index in location s is given by $P_s = \left(\sum_q [p_s(q)]^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}$.

For the numerical exploration of the model, we assume that space is characterized by an evenly spaced grid dividing the unit square into $\overline{N}^2$ smaller squares (i.e., an $\overline{N} \times \overline{N}$ grid). The center of each square is a location. Plants can be set up at all locations. In our example, $\overline{N}^2 = 36$. This results in more than

[46] It is straightforward to show how we recover our specification when θ diverges to infinity,

$$\lim_{\theta \to \infty} \left[\sum_{o \in O} \left(\frac{B_o Z}{w T_{os}}\right)^{\theta}\right]^{\frac{\varepsilon-1}{\theta}} = \lim_{\theta \to \infty} \left[\sum_{o \in O} \left(\left(\frac{B_o Z}{w T_{os}}\right)^{\varepsilon - 1}\right)^{\frac{\theta}{\varepsilon - 1}}\right]^{\frac{\varepsilon-1}{\theta}} = \max_{o \in O} \left(\frac{B_o Z}{w T_{os}}\right)^{\varepsilon - 1}.$$

68 billion (2^{36}) possible permutations of the allocation of plants across space. We have been unable to consistently solve numerically cases (for any parameterization) with a larger set of potential locations. With more locations, locations get closer to each other and thus are "more similar to each other"— given that fundamentals are drawn from continuous distributions. Hence, the pruning approach in Arkolakis et al. (2017) cannot easily eliminate dominated plant configurations, and thus one must resort to an approach that considers all of the potential configurations, which becomes quickly infeasible as the number of locations expands.

We parameterize the model as follows. We let $\varepsilon = 2$, $\theta = 2$, $w = 1$, $B_o = 1 \; \forall \, o$, $R_s = \frac{3}{2}\sin((3\pi x)(\pi y)) + \frac{3}{2}$, $D_s = -\frac{3}{2}\sin((3\pi x)(\pi y)) + \frac{3}{2}$. Further, $T_{os} = \frac{1+\delta_{os}}{\Delta}$, $Z(q, N) = \frac{q}{1+(\Delta^2 N)^{0.45}}$, and $\xi = \Delta^2$. Finally, we solve the plant location problem for 15 firm types, with intrinsic productivity levels $q \in \{3, 4, 5, 7, 9, 10, 12, 13, 15, 17, 18, 20, 22, 23, 25\}$, and we study the resulting plant allocations for $\Delta \in \{9, 1\}$.

Table IX compares the number of plants, $n^*(q)$, chosen by firms of different productivity, q, as we vary Δ. When $\Delta = 9$, all firms set up only one plant, with the exception of firms with the highest productivity levels which set up two plants. Also, all firms place one plant at the same location, the location that minimizes production costs. This follows from the fact that for high Δ, transportation costs are low regardless of distance, so most firms find it optimal to serve customers with just one plant. For $\Delta = 1$ transport costs increase, and firms place substantially more plants across space in order to save on transport costs. Now, we observe substantial variation in the number of plants across firms.

Table IX: Number of plants per firm, across Δ

	q	3	4	5	7	9	10	12	13	15	17	18	20	22	23	25
Δ	9	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
	1	9	9	10	13	15	15	17	18	19	20	20	20	22	22	22

We now replicate the empirical analysis in the main body of the paper. We first divide our $\overline{N} \times \overline{N}$ grid on the unit square into $\mathcal{S}_M = \overline{N}^2/M^2$ sub-squares, each sub-square consisting of an $M \times M$ grid. Within each sub-square, i, let $n_i^*(q)$ be the number of active plants operated by a firm with productivity q, where $\sum_{i=1}^{\mathcal{S}_M} n_i^*(q) = N^*(q)$ denotes the total number of active plants for that firm across all locations. Also, let $\overline{\mathcal{L}}_i$ represent average population in sub-square i, $\overline{\mathcal{L}}_i = (1/M^2)\sum_{s\in i} \frac{D_s P_s^{-\varepsilon}}{w}$.[47] As in the main body of the paper, we use these objects to construct a measure of average density across locations for a firm with productivity q as,

$$\overline{L}(q) = \sum_{i=1}^{\mathcal{S}_M} \frac{n_i^*(q)}{N^*(q)} \times \overline{\mathcal{L}}_i \; .$$

[47] We have that demand at location s satisfies $D_s = wL_s P_s^{\varepsilon}$, which implies that population at location s is given by $L_s = \frac{D_s P_s^{-\varepsilon}}{w}$.

Finally, we define total firm employment as $L(q) = \sum_s l_s(q)$, where $l_s(q)$ is the employment of an active plant of a firm with productivity q at location s.

Table X: Sorting: Firm Size and Local Density in a solved example

	$\Delta = 9$	$\Delta = 9$	$\Delta = 1$	$\Delta = 1$
	(1) $\ln \bar{L}(q)$	(2) $\ln \bar{L}(q)$	(3) $\ln \bar{L}(q)$	(4) $\ln \bar{L}(q)$
$\ln L(q)$	0.0150	-0.0343	0.0167***	0.0172**
	(0.0093)	(0.0215)	(0.005)	(0.0048)
Observations	15	15	15	15
R-squared	0.180	0.180	0.429	0.473
M	2	3	2	3

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

Table X presents the analogue of Table I for the two discrete economies with $\Delta = 9$ and $\Delta = 1$. When $\Delta = 9$, we find no significant relationship between a firm's employment and local weighted employment. When $M = 3$, there is a negative but statistically insignificant relationship between a firm's employment and its local weighted population. In other words, more productive firms seem to have a larger footprint in less profitable locations. This last result showcases how, for high values of Δ, the predictions implied by Proposition 5 can fail to explain the behavior of firms. In contrast, when $\Delta = 1$ the estimated relationship is positive and significant for both values of M. Thus, for $\Delta = 1$, the results are consistent with the predictions of the proposition, and our empirical findings in Table I.

Of course, we acknowledge that this is just an example and that, in other cases, we might need to use even lower values of Δ to obtain allocations in the discrete economy that exhibit the properties derived for our limit economy. Unfortunately, the inability to reliably solve numerical examples with more locations restricts our ability to test the accuracy of our continuous limit in more complex examples. In any case, using our insights for environments where firms have very few plants, seems unwarranted.