Geographic Concentration and Establishment Size: Analysis in an Alternative Economic Geography Model

Thomas J. Holmes* and John J. Stevens†

This version: August 2003

Previous version: November 2002

Abstract

Big cities specialize in services rather than manufacturing. Big-city establishments in services are larger than the national average, whereas those in manufacturing are smaller. We propose an explanation of these and other related facts. The theory is developed in an economic geography model that is an alternative to the standard Dixit-Stiglitz structure. In our tractable structure, which has potentially wider application, firms have monopoly power in local markets but are price takers in export markets.

Keywords: geographic concentration, establishment size, transportation costs, new economic geography

JEL classifications: L11, F10, R10, R30

*Corresponding author at: Department of Economics, University of Minnesota, Minneapolis, MN 55455, USA and Federal Reserve Bank of Minneapolis, Minneapolis, MN 55401, USA.

email <holmes@econ.umn.edu>

†Board of Governors of the Federal Reserve System, Washington D.C. 20551, USA.

email <john.j.stevens@frb.gov>

Note: Holmes acknowledges support from the NSF through grant SES 9906087. Any views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve Board. The authors thank the editor and three anonymous referees for helpful comments.

1 Introduction

The new economic geography has had a substantial impact on our understanding of geographic specialization patterns but little impact on our understanding of plant-level observations. This lack of impact is a surprise, given that the new economic geography, at its core, is about plant-level scale economies. We expect that this theory would have much to say about regional variation in plant size. For example, this theory might help explain why plants in locations that specialize in an industry are substantially larger than other plants in the same industry (Holmes and Stevens, 2002a). This theory is in sharp contrast with other spatial theories, such as those emphasizing knowledge spillovers and externalities, which arguably have no content for plant-level phenomena.

We use the ideas from the new economic geography to begin to integrate the study of regional specialization and the determinants of plant size. To explain the contribution, we divide the remainder of the introduction into two parts. The first gives facts about cities and establishment size. It shows that a strong relationship exists between the specialization patterns of cities and establishment size. It then proposes a parsimonious explanation that simultaneously accounts for all the facts. The second part is an overview of the theoretical model that formalizes our explanation. The model follows the spirit of the new economic geography literature, but it departs from this literature by not using the standard Dixit-Stiglitz structure. The rest of the overview explains our choice of this alternative structure and argues that it may have broader applicability.

1.1 Facts and an Explanation

It is well known that cities of different sizes specialize in different types of economic activities. Table 1 shows industry specialization patterns for United States cities with data from the 1997 Economic Census (U.S. Department of Commerce, 2001). The 328 primary metropolitan statistical areas (PMSAs) are divided into three city sizes: small (PMSA population under half a million), medium (PMSA of half a million to two million), and large (PMSA over two million). The table reports location quotients (LQ) for major industry sectors for the three size classes. The LQ equals the share of industry sales for establishments

located in the given city size class, divided by the share of population in that size class.¹ The table shows that the service industries—wholesale trade, finance and insurance, and professional, scientific, and technical services—are heavily concentrated in the large cities (the LQ increases from approximately 0.5 in the smallest cities to approximately 1.3 in the largest cities for the three service sectors). Manufacturing activity is concentrated in small cities (the LQ declines from 1.15 to 0.86). Retail is spread evenly across city size types (the LQ is approximately 1 in each case). These specialization patterns are well known.²

[Insert Table 1 here]

Less well known is the link between specialization patterns and establishment size. Table 1 also reports establishment size using *Size Quotients (SQ)*. As in Holmes and Stevens (2002a), we define a size quotient in sector i and location j to be the ratio of mean establishment size in sector i at location j to the mean size in sector i in the United States. (We use sales as our measure of size, just as we use sales to calculate the LQ). For example, the manufacturing SQ for small cities is 1.25. This value means that the average manufacturing plant size in small cities is 25 percent larger than the average size in the United States. A clear pattern emerges in this table: Manufacturing establishment size declines as city size increases, but establishment size increases as city size increases for the three service sectors. Retail establishment size also increases as city size increases, but at a much less pronounced rate. In Holmes and Stevens (2002a), we further document the strong connection between specialization and establishment size.

¹Major sectors are defined to include the North American Industry Classification System (NAICS) major sectors with more than \$500 billion in receipts. For the Finance and Insurance sector (NAICS 52), payroll is used instead of sales because PMSA-level sales for sector 52 are not released. For the Wholesale Trade sector and the Finance and Insurance sector, the Census withheld data for a few PMSAs. But the coverage of disclosed data was greater than or equal to 95 percent of all establishments for all the reported figures.

²The patterns regarding services and retail are obvious and long-standing. The pattern involving manufacturing is a more recent phenomenon. Early in the twentieth century, large cities were centers of manufacturing. By the end of the century, however, manufacturing had shifted to smaller cities. Mills and Hamilton (1994) note that a “dramatic shift of manufacturing from large to smaller MSAs” occurred over this period.

We begin our explanation of these facts by invoking the fundamental elements of the economic geography literature—transportation costs and scale economies.³ Transportation costs are particularly high in the service sector. Face-to-face communication is often crucial with services, and the transportation costs of moving people surely dwarf the cost of moving goods. Service-sector activity concentrates in large cities because large home markets make it possible to both economize on the cost of moving people and to achieve economies of scale. The manufacturing sector concentrates in small cities. With the relatively low cost of shipping goods rather than moving people, manufacturing plants in small cities can obtain scale economies by shipping to a national market. The retail sector is spread evenly across city sizes because transportation costs are prohibitively high, and retail services are not traded.

To these basic ideas from the new economic geography literature, we introduce a small perturbation. We observe that in any sector, or even in any broadly defined industry, some subsets of products within these industries are likely to be difficult, if not impossible, to trade, even if most products in the industry can be traded. In almost any industry, some goods must be custom-made rather than standardized, and it is often convenient to have the custom-made articles produced near their point of sale. This convenience is likely at work for clothing, furniture, printing, and so on. In addition, in many industries, having the final stage of production near the final point of sale is important because some products may be difficult to ship after final assembly. Our explanation for why manufacturing plants are small in large cities is simply that such plants tend to engage only in custom work, whereas plants in small cities are large because such plants produce the more standardized products that can be exported to other cities. Analogously, service-sector establishments are small in small cities because they tend to do only the custom work that cannot be traded. The retail sector is different from the manufacturing and service sectors because all retail establishments sell only to local markets. Retail establishments are larger in bigger cities

³Our focus on transportation costs and scale economies ignores standard Heckscher-Ohlin (HO) factors. These factors certainly contribute to geographic specialization, but we believe that HO factors are only a partial explanation. Recent work, such as that of Hanson and Slaughter (2002) and Holmes (2002), supports our claim.

because of the larger local market base, but the size difference between large-city and small-city establishments is not as great as it is for services. Unlike the retail sector, large-city establishments in the service sector are exporters, so the relative size difference is accentuated between establishments in large and small cities.

Our simple explanation for the facts is consistent with an additional fact documented by Bernard and Jensen (1995, 1999). They compared manufacturing plants that export to other nations, not cities, with plants that do not export and found that exporters are substantially larger than non-exporters. One explanation in the trade literature is that there may be a *fixed-cost* component to exporting (see Melitz, 2002). Our explanation is simply that these small manufacturing plants exist solely to provide custom activities that cannot be traded.

1.2 Our Formalization

We formalize our explanation of these facts in a model that is a radical departure from the new economic geography literature based on the standard Dixit-Stiglitz (1977) structure (see Fujita et al. (1999) and Fujita and Thisse (2002), for a comprehensive reviews of this literature). In our model, there is a continuum of locations. At each location there is at most a single producer of any one good, and that producer has a monopoly in the local market. In export markets, however, the firm competes with “foreign” producers (i.e., from other cities) of the identical product, and so export markets are perfectly competitive. In this model of monopolistic competition, a given firm has a monopoly in one market (the local market) and is competitive in a second market (the export market). There is also free entry.

In our parsimonious model, all products in the economy have an identical production function that includes a region of increasing returns to scale. Products differ only in transportation cost. There are products with generally low transportation costs (we call them manufactured goods), products with generally high transportation costs (services), and products with infinite transportation costs (retail). In addition, for both manufacturing and services, we assume that some component of demand can be satisfied only by local production (that is, some component must be custom-made). Locations (cities) differ in population

size. The equilibrium pattern of specialization in this economy and the size distribution of plants follows the explanation outlined previously.

We were led to depart from the standard Dixit-Stiglitz structure because it is ill-suited for analyzing establishment size. As is well known, all of the action in the Dixit-Stiglitz model is on the extensive margin—the addition or subtraction of new firms (See Holmes, 1999, for a further discussion of this point). The equilibrium size of firms is fixed; it is, in essence, a parameter of the model. Our framework allows for a relatively general cost structure and for the possibility of potentially rich variations in firm size.

Besides its suitability for our purpose, our alternative structure may be of broader interest.⁴ We are currently employing a variant of this model in another application (Holmes and Stevens, 2002b). At the margin, the firms in our model behave competitively because they are competitive in export markets. As a result, the definition and construction of equilibrium in our model has properties that are similar in many ways to standard models of perfect competition. This feature is useful, as models of perfect competition tend to be much more tractable and allow for greater generality than do standard Dixit-Stiglitz models. Our structure may be justified on empirical grounds. In our model, the demand elasticity for local sales is strictly less than that for export sales (which is infinity). It seems empirically plausible to us that, for most firms, the elasticity of local demand would be less than that for exports. This claim is certainly consistent with anecdotal evidence that firms often set lower margins in export markets than in local markets (these situations often lead to complaints about “dumping”). In the standard Dixit-Stiglitz model, a crucial element of the framework is that the elasticity be the same in both the local and the export markets. Recent applications of the Dixit-Stiglitz structure, such as Melitz (2002), have gone very far in generalizing the original framework to allow for rich firm heterogeneity. But these new-generation models still retain the property that local and export demand have the same elasticity. In that respect, our structure is quite different from a new-generation model such as Melitz (2002).⁵

⁴In a recent paper, Ottaviano et al. (2002) also provide an alternative to the standard Dixit-Stiglitz structure, one that is very different from the one considered here.

⁵Another way in which our structure differs from Melitz (2002) is that Melitz employs the standard fixed-cost, constant marginal-cost technology, whereas we have a more flexible cost function.

A basic finding of our model—that the medium-transportation-cost sector concentrates in the big cities, while the low-transportation-cost sector concentrates in the small cities—is well established in the literature. Indeed, it is a main point in the original Krugman papers (1980, 1991). It is most closely related to a finding in Amiti (1998), where transportation cost varies across industries, but technology and product differentiation are the same. (This finding contrasts with the original Krugman analysis, where the low-transportation-cost sector has a different technology (constant returns) and different preferences (no product differentiation) compared with the high-transportation-cost sector.) Our analysis is different from Amiti’s in that we introduce nontraded components of demand in all sectors and are concerned with the scale of establishments. Beyond that, the differences are that we have a continuum of locations rather than two locations and that we use our new structure, whereas Amiti works in the standard Dixit-Stiglitz structure. An appealing feature of our continuum analysis is the way in which it highlights the intuition for our specialization result. We derive for each sector an object that is analogous to a bid-rent function from urban economics. These bid-rent functions pin down the specialization patterns across locations in the same way that these functions pin down land-use patterns in urban economic models.

2 A Simple Model to Get Started

The full model we present in the next section is complicated and very different from the usual models of monopolistic competition. Much is going on in the model: Industries differ in transportation costs, locations differ in size, and within each industry a component of demand exists that can be satisfied only by local production. In this section, we present the simplest version of our model with none of the complications for the expositional purpose of showing how the full model works.

There is a continuum of goods indexed by $x \in [0, 1]$. The representative consumer has a Cobb-Douglas utility function

$$U = \int_0^1 \ln q(x) dx, \tag{1}$$

where $q(x)$ denotes the consumption of good x .

The technology for production is the same across all goods. Labor is the only factor of

production. We let $c(q)$ denote the labor input required to produce q units of a particular good and define the average cost (in labor units) as $a(q) = c(q)/q$. We assume that there is a range of output with increasing returns to scale—specifically, the average cost is minimized at a point $q^* \in (0, \infty)$, where q^* is defined as the lowest point at which the minimum is attained (i.e., q^* is the minimum efficient scale). The average cost function can be either U-shaped or can flatten out for $q > q^*$ so that average cost is constant beyond q^* . Assume that for $q < q^*$, marginal cost $c'(q)$ is strictly increasing. Define $a^* \equiv a(q^*)$. A parameterization that we will sometimes refer to is

$$c(q) = q^\alpha + \phi, \tag{2}$$

for $\alpha > 1$. Here the variable cost is q^α , the fixed cost is ϕ , and the average cost is U-shaped with the minimum at $q^* = \left(\frac{\phi}{\alpha-1}\right)^{\frac{1}{\alpha}}$. Figure 1 illustrates the average cost and marginal cost for this case.⁶

[Insert Figure 1 here]

There is a continuum of identical locations indexed by $i \in [0, 1]$. Each location has a population of ℓ workers, each endowed with one unit of labor. We make assumptions below that require ℓ to be small relative to the degree of scale economies.

While labor is immobile, goods can be traded. The transportation cost of shipping a good from one location to another is τ . The transportation cost takes the standard “iceberg” form: In order to move one unit of a good to another location, $1 + \tau$ units must be shipped, so τ units get lost in transit.

Finally, we assume a strategic structure. Firms take prices in other locations as given and behave as perfect competitors in the export market. In the local market, the industry is oligopolistic and firms are assumed to compete in Bertrand fashion.

An equilibrium in this economy takes the following form. At each location the number, or measure, of products produced locally is $n < 1$. The location imports the remaining $1 - n$ goods, paying for the imports with exports of the n locally produced goods. The export market for any particular good is perfectly competitive, and the price a producer receives

⁶Figure 1 is drawn using $\alpha = 2$ and $\phi = 1/2$.

per unit of exports is p^E . A consumer at any location can then import any good at a price of $p^L = p^E(1 + t)$, the competitive export price with transportation costs added.

The combination of scale economies, a small market size at each location, and Bertrand competition, together imply that there is at most one producer of any particular good at each location. That producer sets a limit price of $p^L = p^E(1 + \tau)$ in the local market that exactly matches the import price. In contrast, in the export market for any particular good x , there is a continuum of producers because (1) each location produces a positive measure $n \in (0, 1)$ of products, (2) the set of goods is bounded—the unit interval, and (3) there is a continuum of locations. Therefore, in export markets the producer is a price taker. Note the differences here from Dixit-Stiglitz, in which there is at most a single producer of any differentiated good throughout the entire economy. Here there is a continuum of producers of each good in the entire economy.

At each location entry is free, so profits must be zero in free-entry equilibrium. Thus, average revenue per unit produced must equal average total cost. Since the price p^L for local consumers exceeds the export price p^E , it is immediate that p^L must be above average total cost and p^E must be below average total cost. Although the export price is below *average cost*, the firm nonetheless gains because exports cover *marginal cost*. Indeed, a producer expands its exports to exactly the point at which the export price equals marginal cost.

We now turn to the formal derivation. The export price equals marginal cost of the last unit produced,

$$p^E = wc'(q), \tag{3}$$

where w is the wage rate and total production, q , is the sum of production for local consumption q^L and exported production q^E . Income at a particular location is just the wage income $w\ell$ because profits are zero. Utility is Cobb-Douglas, and so consumers allocate an equal share of income to each good. Since there is a unit measure of goods, spending per good equals $w\ell$, and since local producers match the import price, all goods have the same local price $p^L = p^E(1 + \tau)$. Therefore, the local consumption of each good q^L is the spending

on the good divided by price:

$$\begin{aligned} q^L &= \frac{w\ell}{p^E(1+\tau)} \\ &= \frac{\ell}{c'(q)(1+\tau)}. \end{aligned} \tag{4}$$

Let f denote profits normalized by the wage rate,

$$\begin{aligned} f(q) &= \frac{p^L q^L + p^E q^E - wc(q)}{w} \\ &= \ell + c'(q)(q - q^L) - c(q) \\ &= c'(q)q - c(q) + \frac{\tau}{1+\tau}\ell. \end{aligned} \tag{5}$$

This expression for profit has a straightforward interpretation. The first term is what revenue would be (in labor units) if all consumers paid marginal cost; the second term is total cost (in labor units). These terms together are negative, since in equilibrium, the firm will be operating at a point $q < q^*$, where average total cost is greater than marginal cost. The export customers do pay marginal cost, but local customers pay a markup above marginal cost. The third term is the total amount of markup collected by the firm, and it completes the expression for profit.

Observe that $f'(q) = c''(q)q > 0$, so f is strictly increasing for $q < q^*$. Assume that

$$f(0) < 0. \tag{6}$$

This inequality will be satisfied if ℓ is small enough; for parameterization (2), assumption (6) holds if and only if

$$\ell < \frac{\tau+1}{\tau}\phi. \tag{7}$$

Because marginal cost equals average total cost at this point, $f(q^*) = \frac{\tau}{1+\tau}\ell > 0$. At the unique quantity, $\tilde{q}$, where $f(\tilde{q}) = 0$, the profit on local consumers exactly balances losses from exports, and firms have zero profit. This value of q is the equilibrium quantity in free-entry equilibrium. Figure 2 illustrates how the profit on local consumers offsets the loss on exports.

[Insert Figure 2 here]

The equilibrium number of firms at each location is

$$\tilde{n} = \frac{\ell}{c(\tilde{q})}.$$

Assume

$$\ell < c(\tilde{q}), \tag{8}$$

so that $\tilde{n} < 1$. The assumption is that the number of entering firms at a location is not greater than the number of goods. Even for a very small $\ell > 0$, there will always be a firm producing at the location, as each firm is of measure zero. For parameterization (2), assumption (8) holds if and only if

$$\ell \left(1 + \frac{\tau}{(1 + \tau)} \frac{1}{(\alpha - 1)} \right) < \frac{\alpha}{\alpha - 1} \phi. \tag{9}$$

For small ℓ , both conditions (7) and (9) are satisfied, and the allocation that we have constructed is an equilibrium. Firms maximize profit by extracting the maximum surplus from local consumers, while in the competitive export markets firms set output so that the export price equals marginal cost. Maximized profit is zero, however, so we have a free-entry equilibrium. We have said that each location produces $\tilde{n}$ goods, but we have not specified which goods a particular location produces. This specification is arbitrary, but across the economy as a whole, goods must be apportioned so that each good is produced by the same number of countries. Though, until now, we have not mentioned market clearing in export markets, it must be satisfied because of Walras law and because we have market clearing in all other markets.

3 The Full Model

We now enrich the model so that we can use it to address the issues we raised in the introduction. We extend the model in four ways. First, we allow locations to vary in population size. Second, for each good, we introduce a component of demand that can be satisfied only by local production. Third, we allow goods to differ in transportation cost. Fourth, for each good, we introduce a constant-returns-to-scale backstop technology.

The backstop technology places an upper bound on prices in all sectors; this technology is used for all goods that are not traded, as we limit the extent of scale economies

by assumption. This assumption implies that a firm using the increasing-returns-to-scale technology is “viable” only if it supplies goods to the export market. A firm that uses the increasing-returns technology is competitive in the export market and is a monopolist in its local market. However, the firm’s local market power is limited both by the potential entry of constant-returns competitors and by goods exported from other locations.

3.1 Population

In the full model, locations vary in population size. We let $\ell(i)$ be the population at location $i \in [0, 1]$, and we assume that population is strictly increasing in i . We let $\underline{\ell} = \ell(0)$ be the minimum population and $\bar{\ell} = \ell(1)$ be the maximum population. The maximum population is further restricted by assumption (A2) below, and later by assumption (A3).

By assuming that the population of each location is fixed, we are ignoring the issue of labor mobility. Thus, in its formal structure, our model is more akin to a trade model like Krugman (1980) rather than a regional model with factor mobility like Krugman (1991). In applying our theory to help us understand data across cities, we recognize that there is labor mobility across cities, and therefore view our model as a first cut at determining equilibrium with a continuum of locations. We believe that the forces determining city specialization in our model would remain important if we extended the model to endogenize city size.

3.2 A Local Production Requirement

For each good at each location, we assume that a certain component of the local demand for the good must be satisfied by local production and that a second component may be satisfied either by local production or by imports. Consider, for example, a service good called “grocery wholesaling.” One component of the demand for this good is for the wholesaling of milk, an important service that must be provided locally because of milk’s perishability. A second component, the wholesaling of crackers, could be satisfied by imports because perishability is not an issue. Another example is cabinet manufacturing. One component of the demand for this good is for cabinets that are built into the wall. For such custom work, building the cabinet near where it will be installed will help to coordinate production

better (the model assumes such coordination is a necessity). A second component of the demand is for freestanding cabinets that could be imported from other locations (though with some transportation cost). The model assumes that though these two components differ on the demand side, they are the same on the production side. A grocery wholesaler can provide crackers and milk wholesaling services as part of the same production process; likewise, a cabinetmaker can produce built-in and freestanding cabinets.

We formalize this assumption as follows. Each good is indexed by the pair (x, k) , where $x \in [0, 1]$ and $k \in \{1, 2\}$. We think of the x as indexing the type of good (for example, grocery wholesaling or cabinetmaking) and k as indexing whether the good must be locally produced ($k = 1$) or can be imported ($k = 2$).⁷ Utility is Cobb-Douglas,

$$U = \int_0^1 [\lambda \ln q(x, 1) + (1 - \lambda) \ln q(x, 2)] dx, \quad (10)$$

where $q(x, k)$ denotes the consumption by an individual of commodity (x, k) . In equilibrium, the preference parameter λ will denote the share of income spent on goods that must be locally produced. For $\lambda = 0$, the model is equivalent to the simple model presented in the previous section.

3.3 Transportation Costs

In the full model, goods are allowed to differ in transportation cost. There are three different levels of transportation cost and goods with the same transportation cost are regarded as being in the same sectors. The three sectors are manufacturing, services, and retail. Let μ , σ , and ρ be the fraction of industries in the three sectors, where $\mu + \sigma + \rho = 1$. Assume the goods are ordered such that $x \in [0, \mu)$ are manufactured goods, $x \in [\mu, \mu + \sigma)$ are services, and $x \in [\mu + \sigma, 1]$ are retail goods. Make the extreme assumption that manufactured goods have zero transportation cost, $\tau_M = 0$, retail transportation costs are infinite, $\tau_R = \infty$, and service transportation costs are intermediate, $\tau_S \in (0, \infty)$. Note that goods differ across sectors only in transportation costs; the production technology and the way the goods enter the utility function is otherwise symmetric.

⁷Another way to think about this specification is that consumers exhibit an extreme home bias for $k = 1$ varieties and absolutely no home bias for $k = 2$ varieties.

3.4 Backstop Technology

In the full model, producers have the option to use a second technology that has constant returns to scale. With this alternative technology, γ units of labor are required to produce a single unit of output, independent of the scale of production. This two-technology setup—one increasing returns and the other constant returns—is analogous to the setup in Murphy, Shleifer, Vishny (1989).

We assume that

$$\gamma > a^*. \quad (\text{A1})$$

Thus at a large scale of production, the increasing-returns technology is more efficient. We also assume that for all $\ell \in [\underline{\ell}, \bar{\ell}]$,

$$\gamma < a(\ell/\gamma). \quad (\text{A2})$$

Under this assumption, if a location is in autarky, the backstop technology would be the more efficient technology at the resulting autarky production levels. Only through exports would a producer become sufficiently large to make using the increasing-returns technology economical.

4 Equilibrium

We now characterize equilibrium in the full model. Before we present the formal analysis, we begin with an informal discussion. Our point of departure is equilibrium in the original, simple model.

Consider first what the addition of the retail sector does. This sector has infinite transport costs so its goods are not traded. Since its goods are not traded, assumption (A2) implies that the backstop technology is used everywhere. The utility function is Cobb-Douglas, so we know that in equilibrium, a fraction ρ of the labor force in each location will be employed in retail using the backstop technology, and we can otherwise ignore the retail sector.

Next, consider the joint effect of adding to the model differences in populations across locations and differences in transportation costs across the two traded sectors, manufacturing

and services. The key effect here will be specialization by population size. The bulk of the formal analysis below will work through this interesting pattern of specialization.

Finally, consider the effect of adding the local production requirement through the parameter λ . This addition affects the technical analysis through the zero-profit constraint. In the original, simple model, producers could obtain a markup over marginal cost to local consumers because of the transportation costs. Here, for the λ portion of demand that must be locally provided, producers can obtain an even higher markup because there is no import competition. (The assumption here is that price discrimination is feasible.) Producers are limited in what they can charge for the λ portion of demand because of the backstop technology. Indeed, we added the backstop technology to limit the local producer's monopoly power over the λ portion of demand.

4.1 Definition of Equilibrium

We now provide a formal definition of equilibrium. A producer using the scale technology faces for three markets: the local non-importable ($k = 1$) market, the local importable ($k = 2$) market, and the export market. The quantities produced for these markets are denoted $q^1(\ell)$, $q^2(\ell)$, and $q^E(p^E, \ell)$, where we allow the functions to depend upon the population ℓ of the producer's location. From the producer's perspective, these quantities are perfect substitutes in production, and so

$$q(p^E, \ell) = q^1(\ell) + q^2(\ell) + q^E(p^E, \ell) \quad (11)$$

is the total quantity produced. To conserve notation, for a given location of size ℓ and a given good (x, k) , we define $p^1 \equiv p(x, 1, \ell)$, $p^2 \equiv p(x, 2, \ell)$, $p^E \equiv p^E(x)$, and $w \equiv w(\ell)$, where p^1 and p^2 denote the local prices for the $k = 1, 2$ varieties of good x , p^E is the export price taken as given by firms in each location, and w is the wage rate. Then, given p^E , w , and ℓ , firms maximize the profit function

$$H(p^E, \ell) = p^1 q^1 + p^2 q^2 + q^E p^E - c(q(p^E, \ell))w \quad (12)$$

by choosing a total quantity to produce, $q(p^E, \ell)$, and setting the two local prices, p^1 and p^2 .

An *equilibrium* for this model is a set of price functions $p(x, k, \ell)$, $p^E(x)$, $w(\ell)$, a set of consumption decisions, a set of entry decisions by firms of whether to produce a particular good at a particular location using the scale-economy technology, and output decisions by firms such that the following conditions hold. First, consumers maximize utility given wage income and any share of firm profits (in equilibrium these profits are zero). Second, firms that enter maximize profit. Third, it is not profitable for another firm to enter given the entry and output decisions of other firms. Fourth, supply equals demand for each sector.

4.2 Characterization of Equilibrium

Our main result is summarized by proposition 1 which claims that, besides being unique, the equilibrium in this model is of a particular type, namely, a *zero-profit specialization equilibrium*.

Proposition 1 *There exists a unique equilibrium in this model. This equilibrium is characterized by firms earning zero profits and by each location specializing in a particular sector for its export goods.*

We begin the proof of the proposition by considering *retail goods*. Trade in retail goods is prohibitively expensive, as $\tau_R = \infty$. Therefore, all goods are produced locally by firms using the backstop technology (assumption A2). The normalized local prices for the $k = 1, 2$ varieties of the goods equal the average cost associated with the backstop technology

$$\frac{p^1}{w} = \frac{p^2}{w} = \gamma. \quad (13)$$

The corresponding quantities demanded are

$$q^1 = \frac{\lambda \ell}{\gamma} \quad (14)$$

and

$$q^2 = \frac{(1 - \lambda) \ell}{\gamma}. \quad (15)$$

Retail markets are geographically localized and competitive.

Now consider a representative *service good*. The location must produce the $k = 1$ variety of the good. If the location does not specialize in this good, then p^1 and q^1 for that

good are analogous to the retail case. If the location does specialize in the good, then we must consider the prices and quantities in all three markets. A service firm using the scale technology takes the competitive export price as given and selects the level of output that maximizes profits. We know that the last good sold will be exported, as assumption (A2) ensures that the firm needs to export to survive. Therefore, profit maximization implies output must solve the following first-order necessary condition

$$\frac{p^E}{w} = c'(q), \quad (16)$$

where q denotes the total quantity produced. The firm uses limit pricing in the two local markets. For the $k = 1$ variety, the firm's pricing is limited by the backstop technology, and so the price and corresponding quantity demanded are the same as in equations (13) and (14). In the local market for $k = 2$ service goods, the firm faces competition from imports, and so the limit price is equal to a markup over the export price

$$\frac{p^2}{w} = \frac{(1 + \tau) p^E}{w} \quad (17)$$

or, using equation (16),

$$\frac{p^2}{w} = (1 + \tau) c'(q). \quad (18)$$

The quantity demanded at this price is

$$q^2 = \frac{(1 - \lambda) \ell}{(1 + \tau) c'(q)}. \quad (19)$$

Any product remaining after satisfying local demand is exported,

$$q^E = q - q^1 - q^2. \quad (20)$$

Last, consider a representative *manufacturing good*. This case is analogous to the service good case with $p^E = 1$ and $\tau = 0$. Therefore, p^1 and q^1 are unchanged, but p^2 and q^2 simplify to

$$p^2 = p^E = 1. \quad (21)$$

and

$$q^2 = (1 - \lambda) \ell. \quad (22)$$

All other prices and quantities are the same as in the service good case. Even with zero transportation costs, productive efficiency is not achieved, as the firm retains price-setting power in the local market for non-importable varieties of the good.

Retail profits equal zero. Profits for firms using the scale technology in the other two sectors can be described simultaneously by substituting the appropriate values of p^E and τ . Normalizing the profit function (12) by w and using equations (13) and (14) and (18) to (20), we define profit in labor units as

$$f(\ell, q) \equiv \frac{H(w, p^E, \ell)}{w} \quad (23)$$

$$= \ell + c'(q)q - c'(q)\frac{\lambda\ell}{\gamma} - \frac{(1-\lambda)\ell}{1+\tau} - c(q). \quad (24)$$

From the zero-profit condition $f(\ell, \tilde{q}) = 0$, we find the optimal output $\tilde{q}$ as an implicit function of population size ℓ .

Lemma 1 *For each $\ell \in [\underline{\ell}, \bar{\ell}]$ there exists a unique $\tilde{q}(\ell)$ such that $f(\ell, \tilde{q}) = 0$. The solution $\tilde{q}(\ell)$ satisfies $q^1(\ell) + q^2(\ell) < \tilde{q}(\ell) < q^*$.*

Figure 3 graphically depicts a representative service-sector firm, and is the analog to figure 2 for the simple model. Notice that the role of assumption (A2) is to ensure that the marginal cost and marginal revenue curves cross in the third segment of marginal revenue—the exports segment.

[Insert Figure 3 here]

To determine the equilibrium output level $\tilde{q}$ of a producer at a location of size ℓ , we need not determine the wage w at the location. As we can see in (23), the wage w cancels out in the normalized profit function. Normalizing the profit function by w simplifies the technical analysis—we can determine output size apart from determining wages and prices. This aspect of the model resembles Dixit-Stiglitz, where the equilibrium output of each firm is a simple function of the model's parameters.

Since the function f is increasing in both ℓ and $\tilde{q}$, equilibrium output $\tilde{q}(\ell)$ strictly decreases in the population ℓ of a location. This result is in sharp contrast to what would

happen in a Dixit-Stiglitz setup where output per firm is constant across locations. What is different here is that markup over marginal cost varies over classes of customers, whereas with Dixit-Stiglitz the markup is constant across consumers. Here, the markup over marginal cost is higher on sales to local consumers than it is on exports (where the markup is zero). A firm at a location with a higher population ℓ will necessarily have more local sales. With more profits from local consumers, such a firm breaks even at a smaller overall size $\tilde{q}$. That is why $\tilde{q}(\ell)$ decreases in ℓ .

This is an interesting and surprising finding, but we do not want to emphasize this finding for two reasons. First, in numerical examples that we have considered, the negative effect was quite small. In fact, in the limit where ℓ goes to zero, we can prove that the elasticity of plant size with respect to changes in population ℓ is actually zero. (Remember, small ℓ is in the parameter range we are restricting attention to.) Second, we do not believe that our result about the negative effect of city size is robust to plausible extensions of our basic model. In particular, it is plausible that as markets become larger and more firms in the same industry compete at the same location, the increased competition would tend to drive markups down. Analogous to Breshnahan and Reiss (1991), the lower markups in larger cities would have the effect of increasing the plant size required to break even, which is opposite to the above results. In contrast, the other results that we do emphasize are robust to plausible extensions such as this.

By rearranging equation (16), we find that the equilibrium wage as a function of the location of size ℓ and the sector of specialization (which implies a value for p^E and τ), is

$$w(p^E, \ell) = \frac{p^E}{c'(\tilde{q}(\ell, \tau))}. \quad (25)$$

This function is analogous to the bid-rent function in urban economics (see Mills and Hamilton (1994)), except that we call it a bid-wage function. As in urban economics, the bid-wage function determines what activities occur at what locations. Let $w_S(p^E, \ell)$ denote the bid-wage function for services, and let $w_M(\ell)$ denote the bid-wage function for manufacturing (recall that the manufacturing export price is 1). Figure 4 plots these functions. The equilibrium wage rate increases with location size. This rise occurs because larger locations are relatively more attractive to firms with more price-setting power over a larger local base

of consumers. The activity that supports the highest wage is the activity that the location specializes in.

[Insert Figure 4 here]

Lemma 2 *There exists a unique population cutoff $\hat{\ell}(p^E)$, such that the manufacturing and service wages are equal*

$$\frac{p^E}{c'(\tilde{q}(\hat{\ell}, \tau))} = \frac{1}{c'(\tilde{q}(\hat{\ell}, 0))}. \quad (26)$$

Locations with $\ell \geq \hat{\ell}(p^E)$ specialize in services; locations with $\ell < \hat{\ell}(p^E)$ specialize in manufacturing.

By price discriminating over a larger local base, these service-sector firms have higher profits that, in turn, attract more firms and drive up wages. This result is the classic home market effect (Krugman 1980, 1991). The contribution of Lemma 2 is that the result, though not new, is obtained in a model other than a standard Dixit-Stiglitz or Cournot competition model. For manufacturing, access to low-cost labor is more important than proximity to markets; for services, market access is relatively more important. If the service-sector export price, p^E , were equal to the manufacturing export price of 1, then the graphs of $w_S(p^E, \ell)$ and $w_M(\ell)$ would start at the same point. Because service industries have higher transportation costs, they would always be more profitable than manufacturing industries in the local markets, but equally profitable in export markets. With greater profitability, the service establishments would bid wages up higher than manufacturing, and the graph of the zero-profit wage for services would remain higher than that for manufacturing. In this case, no manufacturing goods would be produced using the scale technology, a situation that cannot occur in equilibrium. For lower service-sector export prices, the graph of the zero-profit wage for services starts below that of manufacturing and crosses the manufacturing graph at a positive level of labor, $\hat{\ell}$. The model is one of agglomeration—the location of labor is fixed. If labor were free to move, then higher wages at larger locations would induce further concentration of labor (absent any other costs such as land rents or longer commute times).

Throughout the preceding discussion, we have implicitly assumed that τ , the transportation cost for services, is not too large. As τ rises, the population cutoff also increases, as only the largest locations can specialize in services. For a sufficiently large τ , the cutoff exceeds the maximum population size, all trade ceases, and each location produces every good using the constant-returns backstop technology. Although this outcome is an equilibrium result, it is not particularly interesting and so we have ignored it throughout the discussion of the model.

To summarize, every location must produce the non-importable ($k = 1$) varieties of manufactures and services (as well as all retail goods). Production of the non-importable goods and services is carried out using the constant-returns-to-scale backstop technology. The production of non-retail importable goods occurs at firms using the scale technology, and each location will have a single firm using the scale technology for any given product. Because locations differ in population, each location will have a measure of firms using the scale technology for distinct goods within the same sector. The determination of which locations specialize in services and which specialize in manufactures is made on the basis of wages, as the immobile workers will work for the firm that offers the highest wage. The tension between market access and low-cost labor makes lower transportation-cost sectors better suited for small locations and high-transportation cost sectors better suited for large locations.

Given export prices and the sector of specialization, firms enter and exit freely, bidding up wages until profits are driven to zero and the labor market clears. The labor supply at location i is simply $\ell(i)$. Labor demand depends upon the measure of firms using the scale technology. Let $n_z(\ell(i), p^E)$ denote the measure of scale technology firms at location i , where $z \in \{M, S\}$ denotes the sector of specialization. In other words, n_z is the measure of the set of x 's produced for export. Labor demand at a location has three components. First, for every non-exported good $x \in [0, 1]$, firms require $q(x, 1, \ell(i))\gamma$ workers to satisfy local demand for the $k = 1$ variety of the good. Second, for every $k = 2$ retail good, $x \in [\mu + \sigma, 1]$, firms require $q(x, 2, \ell(i))\gamma$ workers to satisfy the local demand for that good. Finally, for $x \in [0, \mu + \sigma)$, the location needs $n_z(\ell(i))c(\tilde{q}(p^E, \ell(i), x))$ units of labor if it produces good x . Let $I_{(i,x)}$ be an indicator function that equals one if location i produces

good x using the scale technology and equals zero otherwise. Then, labor market clearing in location i is

$$\begin{aligned}
\ell(i) &= \int_0^{\mu+\sigma} [q(x, 1, \ell(i)) \gamma (1 - I_{(i,x)}) + c(\tilde{q}(x)) I_{(i,x)}] dx \\
&\quad + \int_{\mu+\sigma}^1 [q(x, 1, \ell(i)) \gamma + q(x, 2, \ell(i)) \gamma] dx \\
&= (\mu + \sigma - n_z(i)) \lambda \ell(i) + n_z(i) c(\tilde{q}(x^*)) + (1 - \mu - \sigma) \ell(i) \\
&= \ell(i) [(\mu + \sigma - n_z(i)) \lambda + (1 - \mu - \sigma)] + n_z(i) c(\tilde{q}(x^*))
\end{aligned} \tag{27}$$

where the first integral is manufacturing and service labor demand and the second integral is the retail labor demand. Since each location specializes in a single sector and the price for all goods within a sector are the same, it is without loss of generality that we let x^* represent a typical good produced for export in location i . The second equality then uses the quantity of the representative x produced and the measure of x 's produced using the scale technology to calculate the required labor. Solving for the measure of firms we get

$$n_s(\ell(i), p^E) = \begin{cases} \frac{(1-\lambda)(\mu+\sigma)\ell(i)}{c(\tilde{q}(\ell(i))) - \lambda\ell(i)} & \text{if } \ell(i) \geq \hat{\ell}(p^E) \\ 0 & \text{otherwise} \end{cases}. \tag{28}$$

$$n_m(\ell(i), p^E) = \begin{cases} \frac{(1-\lambda)(\mu+\sigma)\ell(i)}{c(\tilde{q}(\ell(i))) - \lambda\ell(i)} & \text{if } \ell(i) < \hat{\ell}(p^E) \\ 0 & \text{otherwise} \end{cases}. \tag{29}$$

This expression says that the measure of scale sector firms is equal to the total labor allocated to the $k = 2$ varieties of manufactures and services production divided by the total labor demand from scale firms less the labor that, in the absence of a scale sector firm for this good, would be used exclusively for the production of non-importable goods. We assume that $n_s < \sigma$ and $n_m < \mu$ so that there are more goods than firms. This assumption is analogous to assumption (8) in the simple model and it implies a restriction on the set of admissible population sizes,

$$\ell < c(\tilde{q}) \min \left\{ \frac{\mu}{\mu + \sigma(1 - \lambda)}, \frac{\sigma}{\sigma + \mu(1 - \lambda)} \right\} \quad \text{for all } \ell \in [\underline{\ell}, \bar{\ell}]. \tag{A3}$$

The characterization of equilibrium has so far assumed that the export price is constant. We now consider the full equilibrium, taking into account trade between regions and the

determination of the service-sector export price, p^E . We know that $p^E < 1$, as otherwise services would dominate manufactures at all locations, a situation that cannot occur in equilibrium. The bid-wage function for services (see equation 25 and Figure 4) shifts down with decreases in the export price, as lower export prices imply less profit and therefore exert less pressure on wages. Equivalently, lower export prices imply that the firm needs a larger local base on which to price discriminate. Thus, varying the export price will vary the population cutoff. For example, as we lower p^E , more locations will specialize in manufacturing and the supply of services will decrease relative to manufactures. By considering the net supply of services to the export market, we can determine the price needed for market clearing in the service sector. The sum across locations of net service exports should be exactly zero. Market clearing in the manufacturing sector will follow by Walras' law. Let q^1 and q^2 denote the local demands for a representative service good. Then the demand for *all* services at location i is simply the measure of service goods, σ , times the local demand for a representative good,

$$Q_S^D(i) = \sigma [q^1(\ell(i)) + q^2(\ell(i))]. \quad (30)$$

Notice that local demand is independent of the export price. Location i 's total supply of services is

$$Q_S^S(i, p^E) = n_s(\ell(i), p^E) \tilde{q}(\ell(i)) + (\sigma - n_s(\ell(i), p^E)) q^1(\ell(i))$$

where $n_s(\ell(i), p^E)$ is the measure of service-sector firms using the scale technology. The net supply of services (i.e., net exports) by location i is

$$Q_S^S(i, p^E) - Q_S^D(i) = n_s(\ell(i), p^E) [\tilde{q}(\ell(i)) - q^1(\ell(i))] - \sigma q^2(\ell(i)). \quad (31)$$

Recall that $n_s(\ell(i), p^E) = 0$ if location i does not specialize in services. Market clearing requires

$$\int_0^1 [Q_S^S(i, p^E) - Q_S^D(i)] di = 0. \quad (32)$$

This expression can be solved for p^E . The export price affects net supply only through the population cutoff $\hat{\ell}$. A higher value for p^E lowers the cutoff, so more locations specialize in services, thereby raising the net supply.

For $p^E > 1$, all locations will produce services so $n_s = \sigma$ for all i and

$$Q_S^S(i, p^E) - Q_S^D(i) = \sigma [\tilde{q}(\ell(i)) - q^1(\ell(i)) - q^2(\ell(i))] > 0.$$

On the other hand, as p^E gets very small, n_s goes to zero and

$$Q_S^S(i, p^E) - Q_S^D(i) = -\sigma q^2(\ell(i)) < 0.$$

Finally, the left-hand side of equation (32) is continuous and monotonically increasing in p^E .

Therefore, the intermediate value theorem implies that there exists a unique export price.

The proof of Proposition 1 follows directly from Lemmas 1 and 2 and the above construction. Specifically, we have shown that there exists a unique equilibrium in which all firms earn zero profits and each location specializes in a particular sector. The next section discusses the implications of this model economy.

5 Implications of the Model Economy

The following propositions formally state the model's implications and clearly show that the model captures the empirical facts observed in the data.

Proposition 2 (Specialization) *The pattern of specialization over goods that vary only in transportation costs is determined by the size of the populations at the geographic locations.*

Proof. This result is immediate from Lemma 2. ■

Lower transportation-cost goods have a lower threshold size at which production can profitably be commenced using the more productive scale technology. The reason is that relatively more goods arrive at their destination (due to lower iceberg transportation costs). Market access is easier for low-transportation-cost goods, so the location of production matters relatively less. Therefore, low-transportation-cost industries locate in low-wage locations because they have been bid out of the larger high-wage locations by the high-transportation-cost industries, for which proximity to final demand is more crucial. This process of bidding for labor leads to a cutoff level of population, such that locations below the cutoff specialize in low-transportation-cost goods and those above the cutoff specialize in high-transportation-cost goods. Notice that Proposition 2—together with our assumptions on transportation

costs, $\tau_M = 0$, $\tau_R = \infty$, and $0 < \tau_S < \infty$ —produces the same specialization patterns we observe in the data. Specifically, large locations specialize in the high-transportation-cost service industries, small locations specialize in the low-transportation-cost manufacturing industries, and all locations produce the highest transportation-cost goods—retail.

Note again that goods differ only in their costs of transportation; these transportation costs, in turn, determine the pattern of specialization across locations that differ only in population size. We think that differences in transportation cost are a significant way in which industries vary. To highlight these important differences, we have abstracted other ways that industries differ. In reality, we expect industries to also differ in technology and demand characteristics.⁸

Proposition 3 (Establishment Size) *Within an industry, establishments are larger at those locations specializing in that industry.*

Proof. See appendix. ■

Thus, regardless of the industry we are considering, locations with a firm using the scale technology will have larger establishments than locations without a scale establishment, as those with scale establishments will be exporting. With our transportation-cost assumptions, this implies that service-type industries have larger establishments in big cities, while manufacturing industries have larger establishments in small cities and rural areas. Again, this result is borne out in the data. Note that smaller locations have a smaller measure of establishments using the scale technology rather than smaller establishments. Therefore, establishments are still larger in the small location’s sector of specialization. The only way to contradict this result is for a location to be so large that the scale technology can profitably be used to meet local demand without exporting. This possibility is ruled out by assumption A2.

Proposition 4 (Productivity) *Locations that specialize, and hence export, in a given industry exhibit a higher labor productivity (output per worker) than locations not specializing in that industry.*

⁸In Holmes and Stevens (2002b) we consider the opposite extreme where all goods have the same transportation cost, but goods vary in production technology.

Proof. This result follows directly from the two technologies; output per worker is higher when using the scale technology. ■

Since specialization corresponds to exporting in that industry, Propositions 3 and 4 imply that exporting establishments are larger and more productive. This result has empirical support in Bernard and Jensen (1995, 1999), as discussed in the introduction.

6 Conclusion

This paper is an attempt to begin integrating the study of regional-level facts (here specialization) and plant-level facts (here plant size). Our model can account for why large cities specialize in services rather than manufacturing and why the size of establishments in big cities is above the national average in services and below the national average in manufacturing. A crucial ingredient of our theory is that within each sector, some aspects of an industry (for example, custom work) will always need a local presence. Small plants exist to provide this custom work. While this explanation is simple and straightforward, the formal analysis is complex because the firm faces three sources of demand: (1) local demand that imports can meet (non-custom work), (2) local demand that imports cannot meet, and (3) export demand. To sharpen our focus on these sources of demand, we keep the industry side very simple and assume that industries differ only in transportation costs and not in technology or demand characteristics.

If our explanation is on the right track, it suggests that measures of the geographic concentration of industry (for example, Ellison and Glaeser 1997) may in some sense understate the degree of concentration. If we focused only on that portion of demand that can be met by traded goods, industries would be more highly concentrated geographically than if we also included the small plants doing only custom nontradeable work. In Holmes and Stevens (2002a), we find with U.S. data that when we focus on only large plants, the geographic concentration of industries is significantly higher than when we include all plants. This result was obtained using the Ellison and Glaeser index of concentration, which controls for the size distribution of establishments.

In some respects our model is a substantial departure from the usual Dixit-Stiglitz eco-

economic geography model. We believe that our structure—which has firms in a monopolist position in the local market but in a competitive position in the export market—may prove useful in other contexts. It has the benefits of the tractability of competitive models but still captures the insights of the new economic geography literature. In Holmes and Stevens (2002b), we used a variant of this structure to reexamine issues about the home-market effect raised by Davis (1998). Given the large amount of interest that the economic geography literature has generated, and given the well-known limitations of Dixit-Stiglitz models (in terms of their flexibility and generality), an alternative formulation, such as the one discussed in this paper, merits further exploration.

In one important respect, our model is similar to the usual Dixit-Stiglitz economic geography model. This is our model’s crude treatment of geography. The transportation cost is zero for local trades and a positive constant for trade with the rest of the world. This feature of the model underlies our result that the elasticity of demand is finite for local trades and infinite for trades with the rest of the world. A richer geographic structure would allow for a more realistic treatment of distance. We expect that in a model with this richer structure, the elasticity of demand would increase in absolute value with distance in a continuous fashion. We believe our qualitative results about establishment size would continue to hold in such a model, but we have avoided using such a rich geographic structure to keep things simple.

Appendix: Proofs

Note that assumption (A2) implies

$$\ell < c(\ell/\gamma) \tag{33}$$

for all $\ell \in [\underline{\ell}, \bar{\ell}]$. This expression proves useful in some of the proofs that follow.

Proof of Lemma 1. We first establish some properties of f . The function f is strictly

increasing in both ℓ and q ,

$$\begin{aligned}
f_\ell &= \frac{\tau + \lambda}{1 + \tau} - \frac{\lambda}{\gamma} c'(q) \\
&\geq \frac{\tau + \lambda}{1 + \tau} - \frac{\lambda}{\gamma} a^* \\
&> \frac{\tau(1 - \lambda)}{1 + \tau} \quad (\text{assumption A1}) \\
&> 0
\end{aligned}$$

and

$$f_q = c'' \left[q - \frac{\lambda \ell}{\gamma} \right] > 0.$$

Also,

$$\begin{aligned}
f(\ell, q^1 + q^2) &= \ell - c(q^1 + q^2) \\
&< \ell - c(\ell/\gamma) \\
&< 0 \quad (\text{equation 33})
\end{aligned}$$

and

$$\begin{aligned}
f(\ell, q^*) &= \ell + c'(q^*) q^* - c'(q^*) \frac{\lambda \ell}{\gamma} - \frac{(1 - \lambda) \ell}{1 + \tau} - c(q^*) \\
&= \ell \left[\frac{\tau + \lambda}{1 + \tau} - a^* \frac{\lambda}{\gamma} \right] \\
&> \ell \left[\frac{\tau(1 - \lambda)}{1 + \tau} \right] \quad (\text{assumption A1}) \\
&> 0.
\end{aligned}$$

Since f is continuous with $f(\ell, q^1 + q^2) < 0$ and $f(\ell, q^*) > 0$, the existence of a $\tilde{q}$ satisfying the zero-profit condition follows from the intermediate value theorem. Uniqueness of $\tilde{q}$ follows from the monotonicity of f . ■

Proof of Lemma 2. By equating the bid-wage functions for manufacturing and services and rearranging, we get

$$w^S(p^E, \hat{\ell}, \tau) = w^M(1, \hat{\ell}, 0) \quad (34)$$

or

$$\frac{p^E}{c'(\tilde{q}(\hat{\ell}, \tau))} = \frac{1}{c'(\tilde{q}(\hat{\ell}, 0))}. \quad (35)$$

As $\ell \rightarrow 0$, $\tilde{q} \rightarrow q^*$ and so

$$w^S(p^E, \ell, \tau) < w^M(1, \ell, 0), \quad (36)$$

as $p^E < 1$. ($p^E \geq 1$ cannot be an equilibrium price as no manufacturing goods would be produced.) We know that w^S and w^M are both increasing in ℓ as $d\tilde{q}/d\ell < 0$ for fixed τ . Suppose by way of contradiction that as $\ell \rightarrow \bar{\ell}$, the inequality $w^S < w^M$ continues to hold. Then no location would produce services. This statement contradicts p^E being an equilibrium price. Therefore, there exists an $\hat{\ell}$ such that $w^S(p^E, \hat{\ell}, \tau) = w^M(1, \hat{\ell}, 0)$. Uniqueness follows from the fact that

$$\frac{dw^S}{d\ell} = p^E \frac{dw^M}{d\ell} > 0.$$

Locations where $\ell < \hat{\ell}(p^E)$ specialize in manufacturing, as manufacturing firms can outbid service firms for labor. Similarly, locations where $\ell \geq \hat{\ell}(p^E)$ specialize in services. ■

Proof of Proposition 3. Let $\ell_1 > \hat{\ell} > \ell_2$. By Lemma 2 we know that locations 1 and 2 specialize in services and manufactures, respectively. We first show that for a given service good, the employment at the scale technology firm in location 1, $c(\tilde{q}(\ell_1))$, is greater than the employment in all of the constant returns to scale firms in location 2, $\lambda\ell_2$. That is,

$$\begin{aligned} c(\tilde{q}(\ell_1)) &> c(\ell_1/\gamma) \quad (\text{as } \tilde{q}(\ell_1) > \ell_1/\gamma) \\ &> \ell_1 \quad (\text{equation 33}) \\ &> \lambda\ell_2 \quad (\text{as } \ell_1 > \ell_2 \text{ and } \lambda < 1). \end{aligned}$$

We next show that for a given manufacturing good, the employment at the scale technology firm in location 2 is greater than the employment in all of the constant returns to scale firms in location 1. For reasons similar to those in the previous case,

$$\begin{aligned} c(\tilde{q}(\ell_2)) &> c(\tilde{q}(\ell_1)) \\ &> c(\ell_1/\gamma) \\ &> \lambda\ell_1. \end{aligned}$$

■

References

- Amiti, M. (1998) Inter-Industry Trade in Manufactures: Does Country Size Matter? *Journal of International Economics*, 44: 231-255.
- Bernard, A.B., Jensen, J.B. (1999) Exceptional Exporter Performance: Cause, Effect, or Both? *Journal of International Economics*, 47: 1-25.
- Bernard, A.B., Jensen, J.B. (1995) Exporters, Jobs, and Wages in U.S. Manufacturing, 1976-1987. *Brookings Papers on Economic Activity, Microeconomics*, 67-112.
- Bresnahan, T., Reiss, P. (1991) Entry and Competition in Concentrated Markets, *The Journal of Political Economy*, 99: 977-1009.
- Davis, D.R. (1998) The Home Market, Trade, and Industrial Structure. *American Economic Review*, 88: 1264-1276.
- Dixit, A.K., Stiglitz, J.E. (1997) Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67: 297-308.
- Ellison, G., Glaeser, E. (1997) Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105: 889-927.
- Fujita, M., Krugman, P., Venables, A.J. (1999) *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, MA: MIT Press.
- Fujita, M., and Thisse, J.F. (2002) *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*. Cambridge: Cambridge University Press.
- Hanson, G.H., Slaughter, M.J. (2002) Labor-Market Adjustment in Open Economies; Evidence from US States. *Journal of International Economics*, 57: 3-29.
- Holmes, T.J. (1999) Scale of Local Production and City Size. *American Economic Review Papers and Proceedings*, 89: 317-20.
- Holmes, T.J. (2002) The Role of Cities: Evidence from the Placement of Sales Offices. Federal Reserve Bank of Minneapolis, Staff Report #298.

- Holmes, T.J., Stevens, J.J. (2002a) Geographic Concentration and Establishment Scale. *Review of Economics and Statistics*, 84: 682-690.
- Holmes, T.J., Stevens, J.J. (2002b) The Home Market and the Pattern of Trade: Round Three. Federal Reserve Board, FEDS Working Paper #2002-23.
- Krugman, P. (1991) Increasing Returns and Economic Geography. *Journal of Political Economy*, 99: 483-499.
- Krugman, P. (1980) Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review*, 70: 950-959.
- Melitz, M.J. (2002) The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Unpublished manuscript, Cambridge, MA, Harvard University.
- Mills, E.S., Hamilton, B.W. (1994) *Urban Economics*, Fifth Edition. New York: Harper-Collins.
- Murphy, K.M., Shleifer, A., Vishny, R.W. (1989) Industrialization and the Big Push. *Journal of Political Economy*, 97: 1003-1026.
- Ottaviano, G, Tabuchi, T., and Thisse, J.F. (2002), "Agglomeration and Trade Revisited," *International Economic Review*, 43: 409-436
- U.S. Department of Commerce (2001) Economic Census 1997, CD-Rom Disk C1-E97-NA1F-17-US1.

Table 1.
City Specialization and Establishment Size
By City Size and Major Industry

Industry	Location Quotient			Size Quotient		
	Small cities	Medium cities	Large cities	Small cities	Medium cities	Large cities
Service Sectors						
Wholesale Trade	0.54	0.97	1.27	0.70	0.98	1.12
Finance and Insurance	0.52	0.89	1.36	0.54	0.86	1.38
Professional, Scientific, and Technical Services	0.49	0.87	1.40	0.64	0.87	1.23
Manufacturing	1.15	1.06	0.86	1.25	1.08	0.81
Retail	1.01	1.04	0.96	0.91	1.03	1.04

Source: Authors' calculations with geographic data from the 1997 Economic Census (U.S. Department of Commerce, 2001). City size is defined by the 1997 PMSA population: small (under 500,000), medium (500,000 to 2 million), and large (over 2 million). The Location Quotient is the share of sales in the city size class divided by the share of population. The Size Quotient is the average establishment sales size at the location divided by average U.S. sales size.

Figure 1: Technology

Figure 2: Profits and Losses in the Simple Model

Figure 3: A Representative Service-Sector Establishment

Figure 4: Bid-Wage Functions and Specialization