Federal Reserve Bank of Minneapolis
Research Department Staff Report 298

January 2002

The Role of Cities: Evidence From the Placement of Sales Offices

Thomas J. Holmes*

University of Minnesota
and Federal Reserve Bank of Minneapolis

ABSTRACT

What is the force of attraction of cities? Leading explanations include the advantages of a concentrated market and knowledge spillovers. This paper develops a model of firm location decisions in which it is possible to distinguish the importance of the concentrated-market motive from other motives, including knowledge spillovers. A key aspect of the model is that it allows for the firm to choose multiple locations. The theory is applied to study the placement of manufacturing sales offices. The implications of the concentrated-market motive are found to be a salient feature of U.S. Census micro data. The structural parameters of the model are estimated. The concentrated-market motive is found to account for approximately half of the concentration of sales offices in large cities.

*Address: Dept. of Economics, 1035 Heller Hall, University of Minnesota, Minneapolis, MN 55455; Email: holmes@econ.umn.edu. I acknowledge support from NSF grant SES-9906087 for funding this project. I thank the Center for Economic Studies at the U.S. Census Bureau for making it possible for me to access the Census data. I thank Adam Copeland for his research assistance. I am grateful for helpful comments from numerous seminar participants. I particularly want to thank Zvi Eckstein, Gautam Gowrisankaran, Joe Harrington, Shawn Klimek, Sam Kortum, Ariel Pakes, Arnie Resnik, and Jim Schmitz for detailed comments or help. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis, the Federal Reserve System, or the National Science Foundation.

1. Introduction

What is the force of attraction of cities? The economics literature focuses primarily on two theories. The first theory is that cities facilitate the diffusion of knowledge, raising productivity. Recent work on the *knowledge-spillover* theory includes Eaton and Eckstein (1997), Black and Henderson (1999), Glaeser (1999), and Lucas (2001). The second theory highlights the importance of transportation costs, scale economies, and product variety. If a large number of people are concentrated in a single place, by trading amongst themselves they enjoy the consumption of a wide variety of products, without anything (or anybody) having to go a long distance. Krugman (1991) formalizes this result; Fujita, Krugman, and Venables (1999) provide a survey of more recent developments. Though this theory is often called the *economic geography theory,* I call it the *concentrated-market theory,* since the knowledge-spillover theory is also an economic geography theory.[1]

There is little work quantifying the importance of the concentrated-market theory relative to the knowledge-spillover theory.[2] This is an important issue for several reasons. First, the policy ramifications of the two theories are potentially very different. The literature on knowledge spillovers usually models the spillovers as an externality; there is no such externality in the concentrated-market theory. This is not to say that concentrated-market models won't have distortions of their own; the key point is that the structure of the problems is different, making for different welfare analyses. Second, while both theories predict the emergence of cities, they potentially differ in their predictions as to how technological change will affect cities. In particular, there is great interest in understanding how the recent information revolution will affect cities. Third, concentrated-market type models are generally much more complicated to work out than knowledge-spillover models. To motivate

[1] It is common in the literature to add a third theory to this list, *input-market pooling*, but that explanation is beyond the scope of this paper.

[2] Exceptions discussed below include Rosenthal and Strange (2001) and Dumais, Ellison, and Glaeser (forthcoming). There is some work in the international trade literature, Davis and Weinstein (1996, 1999), that examines the related question of distinguishing the concentrated-market theory from the Hecksher-Ohlin theory in explaining specialization of countries.

the use of this more complicated structure, the concentrated-market theory should help us understand empirical observations in a way that the knowledge-spillover theory cannot.

In this paper, I shed light on this broad issue by tackling the following question: What is the relative importance of the concentrated-market factor and the knowledge-spillover factor in determining the location patterns of *manufacturer sales offices*? This is much narrower than my broad issue, but the issues are related. And the narrow question is much easier to answer and has certain properties that make it interesting.

Sales offices are ideal for my study for three reasons. First, as discussed further below, both factors are plausibly important a priori, and this makes for an interesting horse race. Second, sales-office work is in many ways representative of the white-collar, information-oriented work that is now the principal work of cities, so the results may be more broadly applicable. Third, this paper develops a methodology for separating out the concentrated-market factor from other factors and the sales-office sector has certain characteristics that make it ideal for a first use of this methodology.

Distinguishing the concentrated-market explanation from the knowledge-spillover explanation is a challenge. The concentrated-market explanation involves transportation costs and scale economies, and these may be difficult to quantify. It is even harder to measure knowledge spillovers (though Jaffe, Trajtenberg, and Henderson (1993) manage to do this for one case). The recent studies of Rosenthal and Strange (2001) and Dumais, Ellison, and Glaeser (forthcoming) attempt to directly control for these various factors. While these papers do succeed in coming up with some clever measures, the papers recognize that the measures employed have limitations. The new approach developed in this paper avoids these measurement difficulties. It follows the spirit of work in the industrial organization literature that uses demand shifts to infer information about other parameters (e.g., Bresnahan and Reiss (1991)).

To understand the idea of the approach, forget about sales offices for a moment and consider a general industry with differentiated products in which some products have exogenously greater demands than other products. Suppose that the two elements of the concentrated-market factor, transportation cost and scale economies, both matter and that knowledge spillovers do not matter. My key theoretical result is that low-demand and high-demand products will differ in the distribution of production in a systematic way. Products with high demand have sufficient scale to sustain *multiple* locations of production; low-demand products can at best sustain a single plant in the entire economy. This difference in number of plants leads to a distinct difference in the pattern of plant location. As explained below, plants producing low-demand goods are more likely to be found in both the largest cities *and* the smallest cities and are less likely to be found in medium-sized cities, as compared to high-demand products. This implication of the concentrated-market theory is sharply different from that of the knowledge-spillover theory. The latter theory predicts *no* systematic relation between the level of exogenous demand and the pattern of location. The distinct implications of the two theories provides the basis for identification.

In the application of this approach to sales offices, the "product" is the sales service for intermediating a particular manufacturer's product. "Demand" is the total national sales of the particular manufacturer. The first part of the empirical analysis determines how the placement of sales offices varies with the sales size of the firm (the demand measure), using confidential U.S. Census micro data. A salient pattern of these data is that large firms are relatively concentrated in medium-sized cities while small firms are relatively concentrated in both the smallest and the largest cities. This is consistent with the concentrated-market theory and cannot be generated by the knowledge-spillover theory. I interpret this finding as evidence that the concentrated-market theory plays at least some role.

The second part of the empirical analysis estimates the structural parameters of the model. The estimates indicate that the concentrated-market theory is on the order of half of the explanation for why sales offices are concentrated in large cities, with the balance being accounted for by a composition of forces, including knowledge spillovers and Hecksher-Ohlin effects. I conclude that the concentrated-market forces are at least of the same order of magnitude as knowledge-spillover forces.

The third part of my analysis uses the structural estimates to return to the broader issue of why cities form in the first place. The estimated transportation cost savings of consolidating people into large cities are found to be large in that they outweigh a benefit from dispersion, a force that is allowed for in the estimated model. To the extent that the transportation cost factor is of a similar magnitude in other white-collar work, this saving is a force in the formation of cities.

In the remainder of the introduction, I first present some intuition for the main theoretical finding. I then discuss sales offices and further develop why looking at them can shed light on the issues raised above.

A. The Theoretical Result

In the theoretical model, the firm solves a static problem of where to set up sales offices. The environment has four key elements. First, a fixed cost is incurred for each office opened. This scale economy is a force of concentration. Second, a transportation cost is incurred when a city lacks a sales office to intermediate local sales. This is a force of dispersion. Balancing only these two offsetting forces, the firm is led to place offices in the largest cities, where a given expenditure in fixed cost results in the largest savings of transportation cost. These first two forces capture the concentrated-market theory. Third, there is heterogeneity across firms in the relative suitability of different cities as locations for offices. For various idiosyncratic reasons, a particular firm might find Enid, Oklahoma—the smallest U.S. metro area—to be a highly productive place to put a sales office (e.g., the firm might have a manufacturing plant or central headquarters nearby). So this firm may chose to locate in Enid rather than New York City. Fourth, on average, firms tend to be more productive in larger cities than smaller cities. This fourth factor captures the knowledge-spillover force.

If firms were constrained to choose only a single location, there would be no way to distinguish the concentrated-market theory from the knowledge-spillover theory. Firms would tend to locate in the largest cities because of both the transportation cost savings and the productivity advantages. But firms are not constrained to choose a single location. Large firms have sufficient economies of scale to afford multiple office locations, and this opens up the avenue for identification.

My main theoretical result is that when the forces of the concentrated-market theory are at work, the location patterns of large (multi-office) firms will systematically differ from small (single-office) firms in two ways. First, small firms are relatively more concentrated in large cities. A small firm has only a single office and so is likely to put it in New York. Large firms have an office in New York and other very large cities, and they have offices in medium-sized cities like Nashville. Second, and more surprisingly, small firms are more concentrated in the smallest cities as well. To understand the result, suppose that for a particular small firm, the largest cities like New York are unsuitable for various idiosyncratic reasons. So the firm is left to compare medium-sized cities like Nashville with small cities like Enid. Nashville's advantage over Enid in being close to more consumers is relatively inconsequential since the firm sells to the nation as a whole. (Nashville's population is a trivial percentage of the total national population.) So when comparing Nashville and Enid, the small firm puts more weight on other factors besides access to the local consumers. In contrast, a large firm with its multiple offices can obtain relatively complete coverage of medium-sized cities; the objective of close access to customers in medium-sized cities will get relatively large weight in its site location strategy. Another way to think about the intuition is to recall the usual tension between being close to downstream demand or upstream supply in site location decisions (e.g., as discussed Hoover (1975)). Since small firms have only one office, it is not feasible to locate close to demand (except perhaps if they locate in New York).

So they worry about locating close to supply (i.e., a low-cost location, which may be in a small city). Big firms can locate close to demand since they have multiple offices, so they focus relatively more on demand than supply.

B. Why Look at Sales Offices?

Sales offices are the home bases of salespeople who work for manufacturers like General Mills and who make sales calls on wholesalers and large retailers like Wal-Mart. In the 1997 Economic Census (U.S. Bureau of the Census (2001)), there were 29,305 sales offices accounting for approximately $1.3 trillion in sales, $46 billion in payroll, and almost a million employees. According to Census statistics, about half of these employees are directly engaged in selling activities and the other half are engaged in support activities, such as office and clerical work. Sales offices are highly concentrated in large cities. As documented in Section 3, per capita sales of offices are ten times larger in the largest cities than in rural areas.

My first reason for looking at sales offices is that it provides an interesting horse race. There are good a priori reasons to think that both the concentrated-market factor and the knowledge-spillover factor are important reasons sales offices are concentrated in big cities. Consider knowledge spillovers first. A salesperson's job is to match the needs of customers with the products of the firm; information is the essence of this job. A salesperson needs to know the market—not just the product line carried by his or her firm, but also the products offered by competing as well as complementary firms. In a large city, this kind of information is likely to spill over from contacts with others.

There are also compelling reasons to think that the forces at work also include the three factors underlying the concentrated-market theory: transportation cost, product differentiation, and scale economies.

Transportation costs are obviously important for sales offices. Writers in the concentrated-market literature, beginning with Krugman (1991) routinely list manufactured goods as the prototypical good for which transportation costs are important. But surely the transportation cost of moving people dwarfs the cost of moving the vast majority of manufacturing goods. A salesperson in New York may need to participate in a one-hour meeting with a client in Los Angeles. The round trip time cost for such a one-hour meeting is a full working day. Face-to-face communication is clearly a crucial element of success in sales. A priori, it seems very likely that economizing on the transportation cost of moving people is a first-order consideration in determining the spatial patterns of sales-office activity.

Product differentiation in this sector is important; in fact, it is extreme. For most sectors, it can be very difficult to determine the extent of product differentiation. Whether or not Kellogg's Raisin Bran should be considered a differentiated product from Post Raisin Bran is a hard question (see Nevo (2001)). But whatever the answer to this question, it is clear that the intermediation services provided by the sales offices distributing these products are extremely differentiated: It is not possible to obtain the Post product from a Kellogg's sales office, and it is not possible to obtain the Kellogg's product from a Kraft sales office (the corporate parent of Post). The substitution possibilities are zero.

Scale economies also matter. There are obvious fixed costs in setting up sale offices, e.g., the office manager, receptionist, rent.

For these reasons, the concentrated-market theory and the knowledge-spillover theory are both plausible explanations for why sales offices are concentrated in big cities. But there is yet a third explanation that should be brought up at this point. Big cities systematically differ from less urban areas in factor compositions, and the sales-office sector intensively uses factors that tend to be found in big cities. The sector employs very well paid, white-collar workers. (Payroll per person is $50,000 in the 1997 Census.) Workers in large cities on average have higher skill levels than workers in small cities (Glaeser and Mare (2001)). The sales-office sector uses airport services and entertainment services (ballgames, restaurants, for example) intensively, and these services are more readily available in larger cities. Thus sales offices might concentrate in cities for standard Hecksher-Ohlin–type reasons. In the analysis, I am unable to separately identify the productivity advantages arising from knowledge spillover from those arising because of Hecksher-Ohlin effects. However, given my a priori belief that the Hecksher-Ohlin effects are nonnegative, I can obtain a *lower bound* on the concentrated-market effect relative to the knowledge-spillover effect.

My second reason for focusing on sales offices is that the work done by salespeople is in many ways representative of the information-oriented work that is the principal work of cities. This white-collar work includes finance, insurance, legal services, wholesale trade, consultants, and so forth. Much of this kind of work involves people meeting with other people, just like salespeople making calls on consumers. The time cost of travelling to meetings is surely an important consideration.

My third reason for focusing on sales offices is that they are particularly amenable to application of my methodology. My approach examines what happens across products. To make such comparisons, I need clean product boundaries. In general it is difficult to cleanly determine the boundaries of differentiated products and to obtain detailed production data at the level of differentiated products. Census data by SIC codes aggregate differentiated products to a degree that varies from one SIC code to the next. But for sales-office services, the boundaries between products correspond to firm boundaries, and these are cleanly delineated in the Census data. Besides clean product boundaries, I need exogenous shifts in demand across products. Since the costs of the sales offices are only a small percentage of sales, it is plausible that these costs have only a small impact on the quantity of a manufacturer's total sales. Hence, when studying the placement of sales offices, as a first step it might not be unreasonable to assume that total sales are exogenous and use total sales as my measure of demand.

2. The Model

I model the problem of a firm choosing a set of cities in which to put sales offices as well as the allocation of sales across offices. The model is highly stylized, capturing the key tensions in stark form.

The first part of this section describes the environment. The second part presents a formal statement of the firm's problem. The third part determines the solution in the limiting cases where the firm is extremely small and extremely large. The third part also makes the important point that if the firm were constrained to choose only a single office, there would be no way to distinguish the concentrated-market force from other forces.

A. The Environment

There are J cities in the economy, numbered 1 through J. Let $A = \{1, 2, ..., J\}$ denote the set of all cities, and let n_j be the population city j. Cities are ordered by increasing size, $n_j \leq n_{j+1}$. Normalize the total population to equal 1, $\sum_{j=1}^{J} n_j = 1$.

Firms are indexed by i, $i \in \{1, 2, ..., I\}$. Let q_i be the total national sales of firm i. This is taken to be exogenous with respect to the sales-office location decision considered here. Since the expenses of sales offices are a relatively small proportion of total sales (7.5 percent), the assumption of exogenous sales may be a useful approximation for a first cut. Assume, in addition, that the firm's sales are evenly distributed across the cities in the economy in proportion to population; i.e., $q_{ij} = n_j q_i$. This is palatable for consumer goods but is obviously less palatable for products such as agricultural equipment.

The problem faced by firm i is to set up a network of sales offices to minimize intermediation cost, taking the sales quantity $q_{ij} = n_j q_i$ delivered to city j as fixed. There are three components to the intermediation cost: selling costs, trade friction, and fixed costs. I describe each in turn.

Each unit of product sold requires a certain processing called the *selling* activity, and this activity must be undertaken by a facility called a *sales office*. The cost of conducting this selling activity in a particular city by a particular firm is assumed to vary both across cities and across firms. Let c_{ij} be the cost to firm i of conducting the selling activity in city j, per unit of goods processed. I defer until later the important topic of how c_{ij} is determined.

There is a trade friction incurred when a sales office at location j conducts the selling activity for sales at location k. The trade fraction is t_{jk} per unit of sales processed. Assume that

$$
\begin{aligned}
(1) \qquad t_{jk} &= \tau,\, j \neq k, \\
&= 0,\, j = k.
\end{aligned}
$$

Thus the friction is completely avoided if the sales are processed by an office in the same city. If the sales are processed by out-of-town offices, the friction is τ, regardless of which out-of-town office does the selling. This friction is meant to capture the degradation in service when a telephone call is used as a substitute for face-to-face communication. Or it can capture the travel cost to conduct the selling services. But note that this travel cost does not vary with distance between cities.

The final component of the distribution cost is the fixed cost of setting up an office. Assume there is a fixed cost of ϕ that must be incurred for each office set up. This fixed cost is constant across firms i and cities j, unlike the selling cost c_{ij} which does vary with i and j.

I return to the issue of how the selling cost c_{ij} is determined. It equals

$$
(2) \qquad c_{ij} = \bar{c} - \alpha n_j - \beta z_j + \varepsilon_{ij},
$$

for $\alpha \geq 0$ and $\beta \geq 0$. The parameter $\bar{c}$ is a positive constant. The second term is meant to capture the knowledge-spillover effect. The larger the city, the systematically lower the cost on account of spillover benefits. The third term is meant to capture Hecksher-Ohlin factors. Cities vary in a characteristic z_j that can be interpreted, for example, as some measure of worker quality or some measure of airport access or both. To make things as simple as possible, assume the quality characteristic varies in a systematic way with city size. In particular, assume the deterministic relationship $z_j = \omega n_j$, for $\omega \geq 0$. Then the cost can be written as

$$
\begin{aligned}
c_{ij} &= \bar{c} - \alpha n_j - \beta\omega n_j + \varepsilon_{ij} \\
&= \bar{c} - \gamma n_j + \varepsilon_{ij}, \qquad (3)
\end{aligned}
$$

for

$$
\gamma \equiv \alpha + \beta\omega. \qquad (4)
$$

The γ parameter combines the spillover force as well as the Hecksher-Ohlin factor that bigger cities have higher quality inputs. Since both factors come from the supply side, I call γ the *composite supply-side parameter*. From the perspective of the firm's location problem, all that matters is the composite effect. But note that if we were to consider a social planner's problem of choosing how large to make cities, the breakdown (4) would matter. If more people were added to a city but the characteristic z_j of the city were to remain fixed, the effect on cost would be only the spillover effect α and not the composite effect γ. In the data on firm location decisions that I look at, it will only be possible to identify the composite γ and not the breakdown. But since $\beta \geq 0$ and $\omega \geq 0$, a priori, γ is an upper bound for α.

The last term $\varepsilon_{i,j}$ in the cost is a random term that is specific to firm i. This term captures idiosyncratic differences in preferences across firms. For example, it may be that firm i has a factory or a corporate headquarters in a particular city j. If a sales office is located in j, the synergies of having the office near the factory or corporate headquarters reduces the selling cost c_{ij} compared to what it would be otherwise. Or perhaps running the sales office of a particular firm i requires a particular kind of manager or a particular talent that is specific to firm i. This firm-specific talent may be available in some cities and not others.

It is intuitive that the larger the city, the more likely it is that a firm might be able to find an unusual talent. And if its factories were randomly distributed proportionally to population, then the larger the city, the larger the probability of a factory. To capture this intuitive notion, the value of the random term $\varepsilon_{i,j}$ is assumed to be the first-order statistic (from below) of n_j draws of a variable x, i.e.,

$$(5) \qquad \varepsilon_{i,j} = \min\{x_{i,j,1}, x_{i,j,2}, ..., x_{i,j,n_j}\},$$

where the $x_{i,j,k}$ are i.i.d. draws from some fixed distribution $F(x)$.[3] With this cost structure, a city that is twice as large gets twice as many idiosyncratic draws x (e.g., twice the possibility of finding a rare firm-specific talent). A particularly convenient case is where the random $\tilde{x}$ are drawn from the double exponential distribution used in the logit model,

$$(6) \qquad pr(\tilde{x} \geq x) = 1 - F(x) = e^{-\xi e^{x}},$$

with parameters ξ.[4] In this case, the distribution of the first-order statistic remains in the same family,

$$pr(\tilde{\varepsilon}_{i,j} \geq x) = e^{-n_j \xi e^{x}},$$

with parameter $n_j\xi$ instead of ξ.

Observe that with the distribution structure embodied in (5) there are two forces that tend to reduce the selling cost in large cities: (1) the composite supply term $-\gamma n_j$ and (2) the shift to a more favorable distribution of $\varepsilon_{i,j}$ with larger population. Perhaps a more typical approach would include the first force but leave out the second force of the shifted distribution. But that approach would be unsatisfactory for my purposes. I want the parameter γ to capture the forces that would tend to lead firms to locate offices in big cities *disproportionate* to their population; without assumption (5), γ would not capture this. With assumption (5), when $\gamma = 0$ the probability that a given city has the lowest cost is proportionate to population.

[3] There is an abuse of notation here since here n_j is an integer number of draws (corresponding to the location's integer population) whereas elsewhere in the paper n_j indicates a city's population share.

[4] An additional parameter can be added in this specification as a coefficient on x in the exponential function. But without loss of generality, this coefficient is normalized to 1.

B. The Firm's Problem

To formally write down the firm's problem, it is useful to introduce some additional notation. Let χ_j be an indicator variable equal to 1 if the firm opens an office in location j and 0 otherwise. (The i subscript for firm i is implicit for the rest of this section.) Let $\chi = (\chi_1, ..., \chi_J)$ be the vector of indicator variables for each city. Let $y_{j,k}$ denote the fraction of selling services conducted by an office at location j for sales in location k, and let y be the matrix formed by the elements $y_{j,k}$. The firm picks χ, the set of office locations, and y, the allocation of sales across offices, to deliver qn_k sales services to each location. The objective is to minimize the sum of selling cost, out-of-town cost, and fixed costs. Formally, the problem is

$$(7) \qquad \min_{\chi,y} \sum_{j=1}^{J} \sum_{k=1}^{J} y_{j,k} \left(c_j + t_{j,k}\right) qn_k + \left(\sum_{j=1}^{J} \chi_j\right) \phi,$$

subject to the constraint that

$$\begin{aligned} \sum_{j=1}^{J} y_{j,k} &= 1, \text{ for all } k, \\ y_{j,k} &\geq 0, \text{ for all } j,\ k, \end{aligned}$$

and

$$\chi_j = 1, \text{ if } y_{j,k} > 0, \text{ for any } j,\ k.$$

It is straightforward to characterize the solution to this problem. It is useful to define B to be the set of locations with a sales office,

$$B = \left\{j\text{: such that } \chi_j = 1\right\}.$$

If there is a sales office at location j, i.e., $j \in B$, then clearly in any solution this sales office conducts all of the selling activities required for city j, $y_{j,j} = 1$. Suppose in a solution there is some city k without a sales office. The selling service for such a city is conducted by the office with the lowest marginal cost, since the transportation cost is the same for all the out-of-town firms. Denote the lowest cost location as j^*,

$$j^* = \arg\min_{j \in B} c_j.$$

Location j^* is called the *export location* because it handles all out-of-town transactions. In a solution with an export location j^*, if an office is open at $j \neq j^*$, the fixed cost and selling costs for the office at j must be less than the cost of importing the selling services from j^*; i.e.,

$$(8) \qquad c_j n_j q + \phi \leq (c_{j^*} + \tau) n_j q, \text{ for } j \neq j^* \text{ and } j \in B.$$

C. The Solution in Limiting Cases

Observe that the solution to problem (7) depends on the fixed cost ϕ and the firm scale q only through the ratio ϕ/q. In this subsection, I determine the solution in the extreme case where the firm is small (so the ratio ϕ/q is large) and in the extreme case where the firm is large (so the ratio ϕ/q is small). These solutions will be used in the next section to derive the implications of the model. It is also shown here that if firms could choose only a single office, the transportation cost parameter τ and the composite supply effect γ could not be separately identified.

Case 1: The Small Firm

Suppose that

$$(9) \qquad \frac{\phi}{q} > \tau.$$

Under condition (9), there is a single office in the optimal configuration. To see why, recall that the reason for opening up a second office is to reduce the out-of-town cost. But under condition (9), $\phi > \tau q$, so the fixed cost of opening one extra office exceeds the trade friction incurred on the entire population (which again is normalized to equal 1).

It is convenient here to divide the firm's objective function by total sales q and to restate the objective as minimizing average total cost. When the firm has a single office and puts the office at city j, the average total cost is

$$
\begin{aligned}
(10) \quad ATC_j &= c_j + (1 - n_j)\,\tau + \frac{\phi}{q} \\
&= (\bar{c} - \gamma n_j + \varepsilon_{i,j}) + (1 - n_j)\tau + \frac{\phi}{q} \\
&= \left[\bar{c} + \tau + \frac{\phi}{q}\right] - (\gamma + \tau)\, n_j + \varepsilon_j.
\end{aligned}
$$

The first term in the first line is the selling cost per unit. The second term is the trade friction that is incurred on all sales, except for the local sales n_j of the office in j. The third term is average fixed cost. The second line substitutes in equation (2) for c_j. Rearranging terms, we see that the third line expresses average total cost as a constant, a term that depends upon n_j and a random term. The firm picks the location that minimizes ATC_j.

Equation (10) for average total cost highlights the fundamental identification problem faced in this paper. The sum of γ and τ enters multiplicatively with city size n_j. Higher values of τ and γ increase the relative advantage of large cities in the same way. There is no way to separately identify γ from τ. There is no way to sort out the importance of the concentrated-market factor (which depends upon the τ parameter) from the knowledge-spillover and the Hecksher-Ohlin factors (both of which are in the γ parameter).

For the results of the next section, it is useful to derive a formula for the probability that city j is the location with the lowest average total cost. In the logit case where ε_j has the double exponential distribution given by (6), this probability is

$$pr_j = \frac{n_j e^{-(\gamma+\tau)n_j}}{\sum_{k=1}^{J} n_k e^{-(\gamma+\tau)n_k}}.$$

The *location quotient* (LQ) is a commonly used statistic that normalizes sales at a location by population at the location. It equals a location's share of sales divided by a location's share of the population. When the firm is small and the logit case applies, the *expected* location quotient at location j is

$$(11)\quad LQ_j^S = \frac{E\left[\frac{sales_j}{total_j}\right]}{n_j} = \frac{pr_j}{n_j} = \frac{e^{-(\gamma+\tau)n_j}}{\sum_{k=1}^{J} n_k e^{-(\gamma+\tau)n_k}},$$

where the superscript S indicates this is the expected LQ for the "Small" firm. The expected share of sales for location j is simply the probability that the single office is located at j.

Observe that if $\tau = 0$ and $\gamma = 0$, then $LQ_j = 1$ for all j. Here expected sales activity is proportionate to population—there are no τ or γ forces to disproportionately concentrate offices in big cities. But if either $\tau > 0$ or $\gamma > 0$, then LQ_j is strictly increasing in city size n_j.

Case 2: The Large Firm

Now consider the case where ϕ/q is small so that the firm is large. It simplifies things to examine the limiting case where the ratio is zero,

$$\lim_{q\to\infty} \frac{\phi}{q} = 0.$$

Here the only concern in the firm's problem is variable cost. The problem here is a special case of the Eaton and Kortum (forthcoming) model of geography and trade. An office at location j will be the supplier to location k if j has the lowest unit cost at k (including the out-of-town cost if $k \neq j$), i.e., if it minimizes

$$c_j + t_{j,k} = \overline{c} - \gamma n_j + t_{jk} + \varepsilon_{i,j}.$$

In the logit case, the probability this occurs is

$$pr_{j,k} = \frac{n_j e^{\left(\gamma n_j - t_{jk}\right)}}{\sum_{j'=1}^{J} n_{j'} e^{\left(\gamma n_{j'} - t_{j'k}\right)}}.$$

The expected location quotient at j for the large firm is, then,

$$(12)\quad LQ_j^L = \frac{\sum_{k=1}^{J} pr_{j,k}\cdot n_k}{n_j}.$$

3. Implications of the Theory

This section shows how contrasting the behavior of small and large firms can shed light on the force of attraction of large cities. Consider a special case of the model where $\gamma = 0$. Call this the *concentrated-market model* since the other forces have been zeroed out. This section shows that in the concentrated-market model, the location patterns of small and large firms differ in particular ways. An alternative special case is where all the concentrated-market factors are zeroed out; i.e., $\tau = 0$ and $\phi = 0$. In the alternative special case, it is immediate that the firm's location behavior is independent of size. (The scale parameter q factors out of the objective function in (7).) This provides a way to distinguish the concentrated-market model from the alternative extreme model.

For the rest of this section assume that $\gamma = 0$ so that the concentrated-market model applies. For the sake of making a sharp and analytically tractable comparison, this section contrasts the limiting case where the firm is arbitrarily small (the *small firm*) with the limiting case where the firm is arbitrarily large (the *large firm*). Assume that the logit case applies so that the formulae for the expected location quotients derived in the last section hold.

My result depends upon the magnitude of the trade friction τ. In the concentrated-market literature, it is standard to focus on the case where the friction τ is at an intermediate level. When τ is zero, geography is irrelevant; when τ is arbitrarily large, there is autarky. In line with this literature, intermediate τ is the interesting case here, as well. The main result partitions the possible values of τ into three regions, and the intermediate region is the one that is highlighted. I will argue that the concentrated-market model cannot be consistent with the data if τ is very large or if τ is very small.

A. Extreme Values for τ

Since τ is varied in this section, let $LQ_j^S(\tau)$ and $LQ_j^L(\tau)$ be the expected location quotient of city j for the small and large firms. Using formulae (11) and (12), we see immediately that in the limiting case where τ goes to zero, expected sales in city j are proportionate to population for both small and large firms,

$$\lim_{\tau\rightarrow 0} LQ_j^S(\tau) = \lim_{\tau\rightarrow 0} LQ_j^L(\tau) = 1.$$

At the other extreme of τ, the distributions of small and large firms are quite different. Small firms by definition are constrained to open only a single office. Fixing the cost draws of a particular firm, for large enough τ, the home-market advantage of the largest city dominates any difference in selling cost between cities, and so the single office is located in city J. Hence,

$$\begin{aligned} \lim_{\tau\rightarrow\infty} LQ_j^S(\tau) &= 0,\ j < J \\ \lim_{\tau\rightarrow\infty} LQ_j^S(\tau) &= \frac{1}{n_J},\ j = J. \end{aligned}$$

For large firms, when τ is large enough, it is optimal to locate an office in each city, so

$$\lim_{\tau\rightarrow\infty} LQ_j^L(\tau) = 1.$$

Observe that when τ is arbitrarily small and when τ is arbitrarily large, the sales office activity of large firms is evenly distributed across cities in proportion to population. However, in the data, large firm activity is disproportionately concentrated in big cities. Hence, if the concentrated-market model is to be consistent with the data, τ cannot be extremely small or extremely large. This suggests looking at what happens at intermediate values of τ.

B. Intermediate Values of τ and the Main Result

To obtain my result, it is necessary to make an assumption about the distribution of city populations. In the result, there is a difference between medium-sized cities and small and large cities. For this to make any sense, the population distribution obviously has to allow for more than two different city-size types. To put some structure on the population distribution, make the relatively standard assumption that the city-size distribution is Pareto,

$$n_j = n_J \left(1 + J - j\right)^{-\frac{1}{\alpha}},$$

for $\alpha > 0$. In the case where $\alpha = 1$, this reduces to the rank-size rule (or Zipf's law) where city size equals the population of the largest city divided by the city rank. It is a well-known empirical regularity that city sizes tend to obey the rank-size rule. (See Gabaix (1999).)

The formulae for the location quotients are cumbersome, making it difficult to obtain an analytic result. However, they are quite amenable to numerical analysis. With such an analysis, I have obtained the following characterization:

Result. Suppose that $J \geq 10$, and the size distribution is Pareto with coefficient $\alpha \geq 0.44$. Suppose $\gamma = 0$ so that the concentrated-market model applies. Then there exist two cutoff levels of τ, $0 \leq \hat{\tau}_1 < \hat{\tau}_2$, such that

(i) For $\tau < \hat{\tau}_1$, there exists a j', $1 < j' < J$, such that

$$\begin{aligned} LQ_j^S(\tau) &> LQ_j^L,\ j < j', \\ LQ_j^S(\tau) &< LQ_j^L, j > j'. \end{aligned}$$

(ii) For $\tau \in (\hat{\tau}_1, \hat{\tau}_2)$, there exist a j' and a j'', satisfying $1 < j' < j'' < J$, such that

$$\begin{aligned} LQ_j^S(\tau) &> LQ_j^L,\ j < j' \\ LQ_j^S(\tau) &\leq LQ_j^L, j' \leq j \leq j'' \\ LQ_j^S(\tau) &> LQ_j^L,\ j > j''. \end{aligned}$$

(iii) For $\tau > \hat{\tau}_2$, there exists a j'', $1 < j'' < J$, such that

$$\begin{aligned} LQ_j^S(\tau) &< LQ_j^L,\ j < j', \\ LQ_j^S(\tau) &> LQ_j^L, j > j'. \end{aligned}$$

Observe that the condition $\alpha \geq 0.44$ includes the rank-size rule case of $\alpha = 1$. The excluded case of $\alpha < 0.44$ is not empirically relevant because in this case virtually the entire population is concentrated in the largest cities. It is worth noting that the result also holds if the actual size distribution of the 273 cities in the data is used instead of the Pareto (which is not a surprise since the actual distribution is close to Pareto).

Figure 1 illustrates the result for the case where $\alpha = 1$ and $J = 100$. Here the cutoffs are $\hat{\tau}_1 = 2.7$ and $\hat{\tau}_2 = 6.2$. The expected location quotients are plotted as a function of city size for both the small firm and the large firm.

The top panel (A) of Figure 1 illustrates the small τ case, $\tau < \hat{\tau}_1$ (here $\tau = 1$). In this range, LQ^S cuts LQ^L from above, as claimed in the result. Note that in this region, there is little difference between LQ^S and LQ^L; both are approximately equal to 1, consistent with the limit result reported above.

The bottom panel (C) plots the large τ case, $\tau > \hat{\tau}_2$ (here $\tau = 8$). In this range, LQ^S cuts LQ^L from below, as claimed in the result. Note that LQ^L approximately equals 1, consistent with the limit result for large τ. Also consistent with the earlier result, the small firm's sales are heavily concentrated in the largest city. This explains why, for the largest city, LQ^S is so much higher than LQ^L.

The middle panel (B) illustrates the intermediate case where τ is between the two cutoffs (here $\tau = 4$). As stipulated in the result and as illustrated in the figure, LQ^S lies above LQ^L for both small cities and the largest cities. In between the extremes, LQ^L lies above LQ^S. Thus sales of the large firm are concentrated in the medium-sized cities, away from the small and large cities, in comparison to the sales of the small firm.

The intuition for why LQ^S is above LQ^L in the large cities is the same as for the high τ case. But why is LQ^S above LQ^L for the smallest cities? To see why, consider first the LQ^S function for small firms. Notice that aside from the spike near the largest cities, this function is relatively flat. Thus, conditioned upon not locating in the biggest city, the probability of locating in any particular city is roughly proportional to population. This follows because outside of the largest cities, savings in trade cost from locating in a particular city are relatively negligible (since any one city has a small percentage of the population), so the other factors (i.e., the ε_{ij}) make the most difference. Now consider the LQ^L function. The sales of offices in small cities are quite low. This follows because even if the firm opens an office in a small town, the office will not sell to the whole country. (Recall that the large firm will generally have multiple offices.) In contrast, when the small firm opens an office in a small town, this office sells to the whole country. (The small firm has a single office.) Thus sales in small cities are relatively lower for big firms than for small firms.

4. The Location of Sales Offices

The first part of this section documents that sales offices are highly concentrated in big cities. The second part shows that the location patterns, broken down by firm size, follow the pattern predicted by the concentrated-market theory. Details about the construction of tables are relegated to the appendix.

A. Sales Offices Are in Big Cities

Cities are defined to be Census Metropolitan Statistical Areas (MSAs). In the 1997 Economic Census, there are 273 MSAs. The largest is New York, with 20 million people in 1997.[5] The smallest is Enid, Oklahoma, with a population of 57,000. The area outside of MSAs has a population of 54 million people, 20 percent of the total.

Table 1 makes the basic point that sales-office activity is highly concentrated in large cities. The table groups cities (MSAs) by population size and includes a category for non-metropolitan areas. For example, there is a category for cities with a population above 8 million, and this includes three cities, New York, Los Angeles, and Chicago. The first half of the table reports sales-office activity measured on a per capita basis. For each geographic grouping, the activity measure of offices within the geographic category is summed up and

[5] New York is a *consolidated* MSA (CMSA). It is an aggregation of 15 *primary* MSAs (PMSAs), including, for example, the Newark, New Jersey, PMSA, the Danbury, Connecticut, PMSA, as well as, of course, the New York, New York, PMSA. This paper treats a CMSA as a single MSA.

divided by the total population in the geographic category. The bottom half of the table reports location quotients (LQs). Recall that this equals the geographic category's share of national activity measure divided by the geographic category's share of total population.

Table 1 reports that the total sales of offices located in nonmetropolitan areas is $620 per person living in nonmetropolitan areas. Per capita sales increase to $2,700 for small cities (under a half million in population) and to $4,920 for cities in the one-half to two million category. Sales go all the way up to $7,980 and $6,860 for the two largest city-size categories, more than a tenfold increase compared to nonmetro areas. The sales location quotient (LQ) for nonmetro areas is only 0.12. This means that the share of U.S. sales of nonmetro sales offices is only 12 percent of the nonmetro share of population. The LQ increases with city size all the way up to 1.55 and 1.33 for the largest two size classes.

The other measures of sales-office activity reveal a similar pattern. Payroll per capita and operating expenses per capita both increase by a factor of ten, going from the smallest to the largest city-size categories. Employment and inventories also increase, but the factor is five instead of ten.

The significant concentration of office activity in large cities is a recurrent feature of earlier Census years. Table 2 takes the cross section of MSAs and reports the results of a simple regression of the log of sales on the log of MSA population.[6] The population elasticity (the slope of the regression line) for sales ranges from 1.63 in 1987 to 1.71 in 1997. Table 2 also reports a regression of differences in log sales on differences in log population between 1997 and 1982. The estimate from this "fixed-city effect, fixed-time effect" regression is 1.80, which is within the ballpark of the cross section estimates. Thus the pattern that relative sales-office activity increases with size holds within cities over time as well as across cities.

Table 2 also reports the population elasticity when additional city characteristics are included in the regression. These characteristics include a measure of education level of

[6] MSA definitions change from Census year to Census year. This exercise holds fixed the MSA definitions to their 1987 levels. Details are in the appendix.

the workforce, a measure of airport access, and a measure of manufacturing activity (see the appendix for details), all of which we would expect to be associated with higher manufacturer sales-office activity in a city. The additional variables do play some role in the regression, raising the R^2 for 1997 from 0.75 to 0.82, and they lower the population elasticity from 1.71 to 1.56. Still, the population elasticity remains quite high.

One issue that can be raised about this analysis is that the MSA is a crude and sometimes arbitrary definition of a city. The appendix considers a richer geographic analysis that uses county-level data (which is finer than MSA data) and that uses information about neighboring counties. The quantitative results with this more complicated structure are similar to the results with MSA data.

B. Location Patterns Broken Down by Firm Size

The Census data have a firm identifier that makes it possible to link all the offices of the same firm. Define *firm size* to be sales summed up across all the offices of the firm. Table 3 provides some selected statistics for firms in different size categories. Note the clear pattern that larger firms have more offices. The smallest firms (with sales under $25 million) have on average 1.9 sales offices per firm. The largest firms (with sales over $100 million) have 22.5 offices per firm. This is obvious evidence of the extent of scale economies, and this information will be used in the estimation in the next section.

Table 4 shows how the distribution of sales across city-size classes varies with different firm-size categories. The table reports estimated location quotients derived from a logit model that includes dummy variables for industry. The estimated sales shares are determined by evaluating the logit model at the dummy variable means, and they are converted into location quotients by dividing through by population. (See the appendix for details.) The resulting estimates are very close to what one gets with a raw cross tabulation that does not include industry controls.

The pattern predicted by the concentrated-market model with intermediate trade costs is a striking feature of Table 4. Observe first that if we look at the rural areas and the smallest cities, there is a clear pattern that the LQ declines with firm size. In the smallest size class in the rural areas, the LQ is 0.30. This falls all they way down to 0.04 for the largest firm-size class. When we look at the largest city-size class, we see that the location quotients also decline as firm size increases. In the three smallest firm-size categories, the LQs are 1.74, 1.81, and 1.76. But the LQs fall to 1.55, 1.55, and 1.37 for the three largest firm-size categories. Thus for bigger firms, the distribution of sales is shifted away from the very small cities and the very largest cities, toward the medium-sized cities.

The patterns in this table are highly statistically significant in the sense that the hypothesis that the distributions are constant across firm size can be rejected with an extremely high degree of confidence.

The patterns in Table 4 are robust to alternative ways of constructing the table. The table changes little when 1992 data are used instead of 1997 data. The key pattern that the distribution shifts towards the middle as firm size increases continues to hold if establishment counts are used instead of sales. It also holds if *firm size* is defined in terms of numbers of offices rather than total sales.

5. Structural Estimates

The implication of the concentrated-market theory derived in Section 3 was found to be a salient feature of the data in Section 4. An alternative model with only knowledge spillovers or Hecksher-Ohlin factors does not have this implication. This suggests that the concentrated-market factor plays some role. But the concentrated-market factor is not necessarily the only force at work.

As a preliminary attempt to quantify the importance of the concentrated-market factor relative to the other factors, this section estimates the structural parameters of the office-location model. The first subsection describes the estimation procedures. The second subsection discusses the estimates and the goodness of fit of the model. The last two subsections use the estimates to return to the two basic questions of this paper: Why are sales offices in big cities? Why do cities form in the first place?

A. Procedure

For the estimation exercise, I generalize the specification of the selling cost to allow it to depend upon observable city-specific characteristics such as the education level in the city. In particular, the selling cost of firm i in city j incorporates a city-specific term ξ_j,

$$(13) \quad c_{ij} = \bar{c} - \gamma n_j + \xi_j + \varepsilon_{ij}.$$

The variable ξ_j depends upon observable city characteristics,

$$\xi_j \equiv \eta_1 \tilde{z}_{1,j} + \eta_2 \tilde{z}_{2,j} + \eta_3 \tilde{z}_{3,j},$$

where $\tilde{z}_{1,j}$ is the measured city education level, $\tilde{z}_{2,j}$ is airport access and $\tilde{z}_{3,j}$ is the level of manufacturing activity in city j. (These controls were considered in the earlier regression analysis and are defined in the appendix.) Refine the earlier variable z_j to be the skill and amenity levels in a city that are not captured by $\tilde{z}_{1,j}$, $\tilde{z}_{2,j}$, $\tilde{z}_{3,j}$ and continue to assume the deterministic relationship $z_j = \omega n_j$.

Recall that the ε_{ij} random variable is the minimum of n_j draws from an i.i.d. variable x with some distribution $F(x)$. Here I assume x is normal with a mean of zero and variance of σ_x^2. Without loss of generality I can rescale all the costs so that $\sigma_x^2 = 1$. As another normalization, I set the constant term $\bar{c}$ in the cost function to be zero since changing it does not affect any choices.

The parameters that remain to be estimated are ϕ (the fixed cost), τ (the friction), γ (the combined knowledge-spillover and Hecksher-Ohlin parameter), and the coefficients η_1, η_2, η_3 on the city-specific characteristics.

It simplifies computation to discretize firm size. I use six sales size categories, the same ones used in Table 4. Let $h \in \{1,2,3,4,5,6\}$ index these categories. I assume that the per capita sales q_h of each firm in a given category h are equal to the mean per capita sales of firms in the category. Table 5 shows the mean sales per firm in each category and the cell counts.

Offices outside of MSAs are not incorporated in the estimation. These account for less than 2 percent of total sales.[7]

Take a particular value of the parameter vector $\theta = (\phi, \tau, \gamma, \eta_1, \eta_2, \eta_3)$ as given. Take as data the population share n_j of each city j and the city characteristics $(z_{1,j}, z_{2,j,} z_{3,j})$, j from 1 to 273 (the number of MSAs). The problem of a particular firm depends upon its size category q_h as well as its particular draw of the vector $\varepsilon_i = (\varepsilon_{i,1}, \varepsilon_{i,2}, ..., \varepsilon_{i,273})$. The problem cannot be solved analytically, so simulation methods are used instead. The problem faced by the firm is somewhat complex as there are $2^{273} - 1$ different combinations of possible office locations. Simulating the probability of each of these choices can in principle be accomplished by taking random draws of the vector ε_i and solving the firm's problem. However, given the extraordinarily large number of choices, this is not a practical alternative, precluding a simulated maximum likelihood approach.

Instead, I employ a simulated method of moments approach. I focus on matching the aspect of the data that was highlighted in the theoretical section, namely, how the distribution of sales across cities varies with firm size. I also include the moments for the number of offices per firm, as this naturally contains information about the extent of scale economies.

To explain the approach, I define additional notation. Fix the parameter vector θ, the firm size class h, and a given random vector ε. Solve the problem of the firm, and let $s_j^h(\theta, \varepsilon)$ be the share of the firm's total sales originating in an office in city j in the solution. The location quotient at city j for this particular firm is

$$LQ_{,j}^h(\theta, \varepsilon) = \frac{s_j^h(\theta, \varepsilon)}{n_j},$$

[7] *Firm size* is defined as the sum of sales over offices in MSAs.

where, again, n_j is city j's population share. Let $O^h(\theta, \varepsilon)$ be the count of the number of offices the firm opens,

$$O^h(\theta, \varepsilon) = \sum_{j=1}^{273} \chi_j^h(\theta, \varepsilon),$$

where, again, χ_j is an indicator variable for whether an office is opened at location j. Now define the expectations over the random vector ε,

$$\begin{aligned} (14) \quad \overline{LQ}^h_{,j}(\theta) &\equiv E\left[LQ^h_{,j}(\theta, \varepsilon)\right], \\ \overline{O}^h(\theta) &\equiv E\left[O^h(\theta, \varepsilon)\right]. \end{aligned}$$

I am unable to obtain an analytical expression for these expectations. But I was able to use simulation methods to obtain an approximation. I drew the random ε vector 4,000 times and then kept this set of draws fixed. For each size class h, and parameter vector θ, I then solved the firm's problem for each of the 4,000 different random vectors and took the averages to approximate the above expectations. Let $\widetilde{LQ}^h_{,j}(\theta)$ and $\widetilde{O}^h(\theta)$ be the approximation calculated this way. The greater the number of draws, the better the approximation. I stopped at 4,000 because in practice it appeared to be large enough for my purposes.[8]

[8] When I estimate the model using only 1,000 simulations, the results are not that different.

The location quotients are linearly independent, so city 1 is excluded. For a given size class h, the (simulated) moment vector is then

$$(15)\quad \widetilde{m}^h(\theta) = \begin{pmatrix} \widetilde{LQ}_2^h(\theta) \\ \widetilde{LQ}_3^h(\theta) \\ \ldots \\ \widetilde{LQ}_{273}^h(\theta) \\ \widetilde{O}^h(\theta) \end{pmatrix},$$

which is a (273×1) vector. Stacking these across all six size categories yields a (1638×1) moment vector

$$(16)\quad \widetilde{m}(\theta) = \begin{pmatrix} \widetilde{m}^1(\theta) \\ \widetilde{m}^2(\theta) \\ \widetilde{m}^3(\theta) \\ \widetilde{m}^4(\theta) \\ \widetilde{m}^5(\theta) \\ \widetilde{m}^6(\theta) \end{pmatrix}.$$

Now turn to the data. Let NUM^h denote the actual number of firms in size class h from Table 5. Suppose that for each size class h, firms are indexed by i from 1 to NUM^h. Let $LQ_{i,j}^h$ be the location quotient in city j of firm i in size class h, and let O_i^h be the firm's number of offices. The sample analogs to the expectations (14) are

$$\begin{aligned} LQ_j^h &= \frac{\sum_{j=1}^{273} LQ_{i,j}^h}{NUM^h}, \\ O^h &= \frac{\sum_{j=1}^{273} O_i^h}{NUM^h}. \end{aligned}$$

Suppose the sample analogs are stacked in an analogous way as (15) and (16), and let m be the (1638×1) vector of the moments in the data.

The simulated method of moments estimate of θ is obtained by minimizing

$$\min_{\theta}[m - \widetilde{m}(\theta)]' V^{-1}[m - \widetilde{m}(\theta)]. \quad (17)$$

The weighting matrix V^{-1} is an estimate of the optimal weighting matrix obtained by using the sample variance of the firm-level moments. I had problems differentiating $\widetilde{m}(\theta)$ because it is a step function (given the discrete choice nature of the problem and the fact that the number of simulated draws is finite). This precluded the use of a gradient-type method to solve (17). Instead, I used a simplex-type method called the *amoeba method*. A bootstrap procedure was used to approximate standard errors. Details about the bootstrap procedure are contained in the appendix.

B. The Estimates

Table 6 reports two sets of estimates. In the first specification (Model 1), the additional city characteristics are constrained to be zero ($\eta_1 = \eta_2 = \eta_3 = 0$). In the second specification (Model 2), the city characteristic parameters are allowed to be nonzero.

In Model 1, the estimates of τ and γ are approximately the same (1.7 and 1.6). Recall that in the problem of a small firm locating a single office, the sum $\tau + \gamma$ is multiplied by population size n_j to determine the cost of locating at j (equation (10)). Therefore, my estimates show that the τ parameter and the γ parameter make virtually the same contribution in inducing a small, single-office firm to disproportionately locate in large cities. In Model 2, the estimate of τ is stable, but the estimate of γ almost doubles. The standard error of γ is much higher than for τ. (The γ parameter tended to bounce around in other specifications I considered while τ was stable.) Even with the doubling of γ, the τ parameter continues to play a large role in explaining why small, single-office firms disproportionately locate in big cities. Here the contributions of τ and γ are $\frac{1}{3}$ and $\frac{2}{3}$, respectively.

The estimates of the fixed cost parameter ϕ are relatively small. To interpret the magnitude, it helps to know that the population share of New York—the largest U.S. city—is $n_{NY} = 0.1$ and the share of Enid, Oklahoma—the smallest U.S. city—is $n_{Enid} = 0.0003$. The smallest firm type ($h = 1$) has per capita sales of $q_1 = 7.5$ (Table 5). If the smallest firm were to locate an office in New York, the savings in out-of-town costs would be $\tau q_1 n_{NY} = 1.7 \times 7.5 \times 0.1 \approx 1.2$. This savings is approximately the same as the fixed cost in Model 2 and about 50 percent higher than the fixed cost in Model 1. This doesn't mean a small firm will always open an office in New York, because the firm also has to consider the idiosyncratic cost realization ε_{NY} for New York. Note that given the small population of Enid, the fixed cost swamps the out-of-town cost savings from opening up an office there. The largest firm type ($h = 6$) has per capita sales of $q_6 = 4,856.9$. For such a firm, the savings in out-of-town cost from opening in Enid is $\tau q_6 n_{Enid} \approx 2.3$, which is larger than the fixed cost, but not so much larger as to make the fixed cost immaterial. The largest firms will not all necessarily have an office in Enid because the selling cost in Enid also matters, in addition to the out-of-town cost and fixed cost.

In Model 2, the estimates for the airport and manufacturing activity variables are both positive, as expected. The coefficient on the education variable is approximately zero. (It is actually negative in sign.)

The chi-squared statistic for both models is quite high. It is not surprising that a highly stylized model with a few parameters and 3,786 firms would fail a conventional statistical test.

Nonetheless, given all that it is being asked to do, this highly stylized model does a good job of fitting the data. Table 7 compares moments in the actual data (Panel A) with moments in the Model 1 economy (Panel B). (Panel C is discussed in the next subsection.) Model 1 captures the qualitative features of the actual data, whether we look up and down the columns or across the rows. In the actual data, LQ^{sales} declines as we move down the table for the largest and smallest cities, while it increases for cities in the 2 million–8 million category. The directions of these effects are the same in the Model 1 economy. The model underpredicts offices per firm for the smallest size class and overpredicts offices per firm for larger firms. The results for the Model 2 economy are similar, but are not displayed here.

In addition to the two models presented here, I have considered various alternatives. The estimates change little when I use 1992 data instead of 1997 data. The estimates change little when I exclude the smallest firm-size class, which consists of over half of all the firms (but a small portion of total offices). I have also estimated a version of the model with substantially fewer moments, by aggregating the location quotients to the aggregated city-size categories used in Table 7. The estimate of γ falls substantially to 0.07. But the estimates of τ and ϕ are 1.14 and 0.35, which are in the same ballpark as the Model 1 estimates. Thus, my finding that the concentrated-market factor is important is robust to alternative specifications.

C. Why Are Sales Offices in Big Cities?

A goal of this paper is to separate out the importance of the concentrated-market factor in accounting for the concentration of sales offices in big cities. One way to do this is to zero out the other forces of concentration to see how much concentration would result from the concentrated-market factor alone. Panel C in Table 7 takes the estimate of ϕ and γ from Model 1, but sets $\gamma = 0$ so that the knowledge-spillover and Hecksher-Ohlin forces are zeroed out. It is evident in the table that there still remains substantial concentration, even though $\gamma = 0$. For example, in the smallest firm-size class, the LQ in the biggest cities is 1.38 as compared to 1.74 with the original value of γ in Model 1. So concentration is about half as large, when concentration is defined as the difference between the LQ and 1. Interestingly, for the larger firm sizes, the difference in concentration is even less. For example, for firms in the 50 million–100 million size class, the LQ is 1.60 with $\gamma = 0$ compared to 1.81 in the benchmark case, a relatively small difference.

If we were to use the parameters from Model 2 as the benchmark, the differences are somewhat larger since γ is larger in Model 2. Nonetheless, with the exception of the smallest sales-size class, concentration in the largest cities with $\gamma = 0$ is about half of what it is in the benchmark model. I conclude that the concentrated-market factor is approximately half the story for why sales offices concentrate in large cities.

D. Why Form Cities?

The model does not endogenize city formation. The sales-office sector is viewed as small relative to the economy as a whole, and the city sizes $(n_1, n_2, ..., n_J)$ are taken as given. Nonetheless, some sense of the forces that lead to city formation can be obtained by analyzing how total intermediation costs would change if population were concentrated in large cities.

To keep the analysis simple, consider a case of a firm that is so small that it opens a single office. Suppose, hypothetically, the firm's problem were changed in two ways First, the structure of cities changes, with New York absorbing the entire population of all the other cities. Second, the firm is constrained to locate only in New York. The savings in out-of-town cost per unit of sales from this change have a lower bound of $(1 - n_{NY}) \times \tau = 0.9 \times 1.7 \approx 1.5$. If the firm would have located in New York anyway, this is the exact savings. If the firm would have located outside of New York, the savings in trade friction are even greater.

To set a sense of the magnitude of these savings, consider the potential impact of this change on the firm's selling cost. Recall that the selling cost per unit equals

$$(18) \quad c_{ij} = \bar{c} - \gamma n_j + \varepsilon_{ij}.$$

For now, assume that the population shift to New York were to leave c_{ij} and its components fixed for all i and j. If the firm always opens in New York, it may be forgoing a location with a lower selling cost. For now, focus on the random ε_{ij} term. The expected value of the first-order statistic and the expected value of ε_{ij} for New York are

$$(19) \quad \begin{aligned} E\left[\min_j \varepsilon_{i,j}\right] &= -3.6, \\ E[\varepsilon_{i,NY}] &= -2.9. \end{aligned}$$

If a firm were to always locate in the city with the lowest ε_{ij}, its expected selling cost would be

0.7 more units than if it always located in New York. The cost increases through the ε term from being constrained to locate in New York can be no more than 0.7. The lower bound on the savings in trade cost is twice as high as the upper bound on the cost increase from the higher ε. Put another way, the savings in the trade friction from concentrating population in New York substantially outweigh the gains from specializing according to comparative advantage (i.e., having the lowest ε_{ij} city get the office).

Of course, shifting the entire U.S. population to New York should have some effect on the cost c_{ij} at each location to the advantage of New York. Unfortunately, my analysis cannot pin down this effect. Recall that γ has a component due to knowledge spillovers and a component due to Hecksher-Ohlin. My procedure cannot disentangle these separate effects. In the extreme case where all of γ is knowledge spillovers, shifting the entire U.S. population to New York will result in savings from knowledge spillovers that are approximately the same order of magnitude as the savings in out-of-town costs discussed above.

6. Conclusion

The narrow question of this paper is, Why are sales offices in big cities? To answer this question, the paper develops a new theory of the site location decision of a firm with potentially multiple establishments. It shows that when the concentrated-market factor emphasized by Krugman (1991 and elsewhere) are at work, the distribution of sales activity for large firms will be relatively concentrated in medium-sized cities, while for small firms, it will be concentrated in small and large cities. Analysis of U.S. Census micro data shows that this implication is a salient feature of sales offices in the United States. A first cut estimation of the model indicates that the concentrated-market factor is approximately half the explanation for why sales offices are in big cities.

The broader question of this paper is, Why are there cities? The work here can only speak to the benefits of city formation to the sales-office sector, a sector that makes up 1 percent of the national payroll. The preliminary estimates of this paper indicate that the savings in out-of-town costs or trade frictions that accompany the formation of large cities are large compared to any offsetting force of dispersion of comparative advantage. To the extent that broader sectors such as finance and business services are similar to sales-office activity, reductions in trade frictions will be a significant force leading to the formation of cities.

A number of simplifications were made in the analysis to get started. In future work, the analysis could be enriched in many ways. The theory recognizes that firms may have idiosyncratic reasons to locate a sales office in a particular city. By linking the sales-office data with the micro data on plant locations from the Census of Manufacturers, it might be possible to quantify location-specific benefits to particular firms. The theory assumes that all sales are intermediated by a firm-owned sales office. But the analysis could be extended to allow firms to decide between a sales office and a merchant wholesaler. The empirical analysis could be extended to exploit the panel nature of the data set.

The sales-office sector has special attributes that make it uniquely interesting to study. But the theoretical and empirical findings of this analysis can potentially be extended to other sectors. With the increasingly pervasive use of scanners, it may be possible to measure the total industry output levels of narrowly defined differentiated products. To the extent that the concentrated-market factor is important, my findings suggest that differentiated goods with a small national market (a market that supports relatively few plants) will tend to be in large cities and small cities. Differentiated products with a large national market (one that supports a large number of plants) will tend to be in medium-sized cities.

Appendix: Notes for Section 3, 4, and 5

7. Notes for Section 3

A. Table 1

The figures in Table 1 are estimates constructed with the publicly available geographic data. The U.S. Census Bureau is reluctant to release new raw tabulations of geographic data since it might be possible to combine this information with previously released information to back out information that the Census did not intend to release.[9] Disclosure problems do not arise with the econometric estimates reported after Table 1. I have recalculated Table 1 with the raw data, and the differences are immaterial. The coverage of the publicly released data ranges from 95 percent of establishments in the over 8 million city-size category to 62 percent in the under half a million category.

B. Table 2

MSA definitions change from Census year to Census year. In order to use a fixed definition of MSAs for this analysis, I used the 1987 definition of MSAs for Table 2. Outside of New England, this is easy to do because MSAs are defined as aggregations of counties, and county definitions (with minor exceptions) do not change over time. In New England, MSAs are not aggregations of counties. However, for the 1987 Economic Census, the Census Bureau provides county-based definitions of MSAs that approximate the actual MSAs in New England. The 1987 county-equivalent MSA definitions were used to define MSAs for all the regressions in Table 2, except for the regression with additional controls, which used the 1997 Census definition.

The variables used in the regression with additional controls are defined as follows. The education measure is the percentage of workers 25 years and older with a bachelor's,

[9] For example, the public data release includes the sales data for the entire Chicago area except for the seven establishments in the Kankakee PMSA. If I were to make a table that specified total sales in Chicago, then one could back out the sales of the seven establishments in Kankakee that the Census did not intend to disclose.

graduate, or professional degree in the MSA in 1990. The source is the U.S. Bureau of the Census (1996). The airport variable is domestic enplanements in 1999 per person. The source is the U.S. Bureau of Transportation Statistics (2000). The manufacturing intensity measure is sales of manufacturing plants per person. The source is the 1997 Economic Census (U.S. Bureau of the Census (2001)).

Table A1 shows a cross tabulation of these three variables. Larger cities have a higher fraction of college-educated workers and have more enplanements per person. Manufacturing activity tends to be concentrated in small cities rather than large. Table A2 reports the results of running log sales on log population, log education, the level of airport activity, and log manufacturing activity. The elasticity estimate for education is large at 0.30, but it is not statistically significant. The elasticity estimate for manufacturing activity is sizable at 0.62. The magnitude of the airport variable is relatively high, and both the manufacturing and airport coefficients are highly statistically significant.

C. City Definitions

This subsection shows that the results reported in Tables 1 and 2 of the main paper are robust to consideration of an alternative, richer geographic structure.

The MSA can be a murky definition of a city. Half of New Jersey is in the New York consolidated MSA, and the other half is in the Philadelphia MSA. Where to draw the line in New Jersey may be arbitrary, but it has the big effect in the analysis of determining whether a particular part of New Jersey is classified as being in an MSA with 20 million people or one with 6 million people. When this boundary line is crossed, there is a discontinuous increase in population.

The alternative analysis discussed here does not rely on the arbitrary way that MSA boundaries may be defined and reduces the problem of discontinuities. The analysis uses data on the 3,111 counties in the 48 contiguous states (plus the District of Columbia). An advantage of this procedure is that the county is a significantly finer geographic unit than the MSA. (For example, the New York metro area is made up of 30 counties.) Moreover, by using county data, I am able to incorporate non-MSA counties into the statistical analysis as well as MSA counties.

For each county i, I determined the identities of all counties within 30 miles of the same county i.[10] Let $neigpop_i$ denote the total population of these neighboring counties, including the population of county i. Counties were ranked by $neigpop_i$ and divided into 10 regions of counties of approximately equal population. For example, the first region consists of 1,395 counties with the lowest values of $neigpop_i$, and together these sparsely populated counties account for approximately 10 percent of the continental U.S. population. The top group with the highest values of $neigpop_i$ consists of 17 counties and 10 percent of the population. For each of these regions, using the 1997 data, I calculated the various measures of sales-office activity on a per capita basis and made a cross tabulation for the different $neigpop_i$ groupings similar to the first part of Table 1. Disclosure concerns preclude publication of these tables, but it is possible to report summary statistics. Table A3 reports a regression of the (log of the) per capita activity measure on the (log of the) average neighboring population. For example, the coefficient when per capita sales is the activity measure is 0.64. To compare this to the results from the MSA regressions in Table 2, we need to add 1 to the coefficient in Table A3 (since the left side variable here is a per capita number). Adding 1 to 0.64, we see that the estimate of 1.64 is quite close to the earlier MSA regression result of 1.71 for the same year. Table A3 reveals the same patterns as does Table 2. The largest effects are on sales, payroll, and operating expenses, while employment and inventories are in between, but still quite large. The same analysis for the earlier years yields similar results.

[10] The geographic centroid of the county was used to define county location.

8. Notes for Section 4

A. Procedure Used to Construct Table 4

I estimated a logit model for the distribution of the number of offices and the sales of offices for the same MSA size categories as in Table 1. The firm-size categories are (in millions of dollars) under 25, 25–50, 50–100, 100–250, 250–1000, and 1000 and over. Let s denote a particular city-size category ($s = 1$ for non-MSA, $s = 2$ for MSAs with less than half a million, etc.). Let h denote a particular firm-size category ($h = 1$ for under 25 million, $h = 2$ for 25–50, etc.). Suppose that the probability that a particular office i locates in a location of type s depends upon the size class h of the firm that the office is part of as well as the industry k of establishment i. (Note that industry is defined at the establishment level, not the firm level.) To write this in a multinomial logit fashion, let x be a vector of dummy variables for firm size, so that $x_{h,i}^{size} = 1$ if establishment i is in size class h and is zero otherwise. Analogously, suppose that y is a vector of dummy variables for industry, so that $y_{k,i}^{ind} = 1$ if establishment i is in industry k and is zero otherwise. Let α^s and β^s be a vector of coefficients for choice s. Then with the multinomial logit specification, the probability that an office is in s, given h and k, is

$$prob(\text{office in } s|h,k) = \frac{e^{\alpha^s x_i + \beta^s y_i}}{\sum_{s'=1}^{5} e^{\alpha^{s'} x_i + \beta^{s'} y_i}}.$$

I used maximum likelihood to estimate $(\alpha^1, \beta^1, \alpha^2, \beta^2, ..., \beta^5, \beta^5)$. I then used the parameter estimates to calculate the estimated probability of being in a particular size class s, for each given level of h, and given that industries' dummy variables are evaluated at the mean across all establishments (so each is the fraction of offices in the given industry). To make the analysis comparable to the earlier tables, I divided each probability by the fraction of population in the area to create a location quotient measure.

The above procedure calculates an *establishment* location quotient. In Table 4 of the paper, I report a *sales* location quotient. This is constructed in a similar manner as above, except the measure is sales-weighted. In particular, I regarded each million dollars of sales as a separate observation and estimated the model above to determine the probability that the million dollars in sales is allocated to each city-size class. I then calculated the location quotient in the analogous way. The broad patterns in the establishment location quotient (not shown) are the same as for the sales location quotient (Table 4).

9. Notes for Section 5

The additional city characteristics used in Model 2 are specified in levels. The units are defined in Table A1.

Since the objective function is a step function, it proved to be difficult to approximate derivatives. This precluded me from obtaining estimates of the standard errors in the usual way. Instead I used a bootstrap procedure. Let $\hat{\theta}_1$ be the parameter estimates for Model 1 reported in Table 6. Setting $\theta = \hat{\theta}_1$, I drew 25 simulated data sets. (For Model 2, I drew 21 simulated data sets.) Each simulated data set has the same number of firms in each size class as in the actual data. I then reestimated the model with each of the simulated data sets. The reported standard error is the statistic calculated from the distribution of the 25 estimates (21 for Model 2).

When I calculated the weighting matrix for the simulated data in the same way I calculated that matrix for the actual data, the weighting matrix was singular. (This didn't happen with the actual data). My solution was to take $\hat{\theta}$ and approximate the optimal weighting matrix. This was accomplished by simulating the distribution of the moments with a large number of firms of each type (10,000) rather than the actual number of firms. This fixed, optimal weighting matrix was used in the estimation procedures with the simulated data. As a result, reported standard errors do not take into account sampling variation of the weighting matrix. They do take into account sampling variation of the moments as well as the approximation error (since a new set of ε draws are obtained for each estimation set).

In Model 2, it is optimal for the smallest firm to always have a single office, so there is no variance of this moment. Hence, even the optimal weighting matrix is singular here.

For this case, I substituted into the weighting matrix an estimate of the actual variance of this moment in the real data in place of zero.

References

Black, Duncan, and Vernon Henderson (1999), "A Theory of Urban Growth," *Journal of Political Economy* 107 (April): 252–84.

Bresnahan, Timothy F., and Peter C. Reiss (1991), "Entry and Competition in Concentrated Markets," *Journal of Political Economy* 99 (October): 977–1009.

Davis, Donald R., and David E. Weinstein (1996), "Does Economic Geography Matter for International Specialization?" Working Paper 5706, National Bureau of Economic Research.

__________ (1999), "Economic Geography and Regional Production Structure: An Empirical Investigation," *European Economic Review* 43(2): 379–407.

Dumais, Guy; Glenn Ellison; and Edward Glaeser (forthcoming), "Geographic Concentration as a Dynamic Process," *Review of Economics and Statistics.*

Eaton, Jonathan, and Zvi Eckstein (1997), "Cities and Growth: Theory and Evidence From France and Japan," *Regional Science and Urban Economics* 27: 443–74.

Eaton, Jonathan, and Samuel S. Kortum (forthcoming), "Technology, Geography, and Trade," *Econometrica.*

Fujita, Masahisa; Paul Krugman; and Anthony J. Venables (1999), *The Spatial Economy: Cities, Regions, and International Trade*, Cambridge, Mass.: MIT Press.

Gabaix, Xavier (1999), "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics* 114 (August): 739–67.

Glaeser, Edward L. (1999), "Learning in Cities," *Journal of Urban Economics* 46 (September): 254–77.

Glaeser, Edward L., and David C. Mare (2001), "Cities and Skills," *Journal of Labor Economics* 19 (April): 316–42.

Henderson, J. Vernon (1986), "Efficiency of Resource Use and City Size," *Journal of Urban Economics* 19: 47–70.

Hoover, Edgar M. (1975), *An Introduction to Regional Economics,* 2nd ed., New York: Knopf.

Jaffe, Adam B.; Manuel Trajtenberg; and Rebecca Henderson (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108 (August): 557–98.

Krugman, Paul (1991), "Increasing Returns and Economic Geography," *Journal of Political Economy* 99 (June): 483–99.

Lucas, Robert E., Jr. (2001), "Externalities and Cities," *Review of Economic Dynamics* 4 (April): 245–74.

Nevo, Aviv (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica* 69 (March): 307–42.

Rosenthal, Stuart S., and William C. Strange (2001), "The Determinants of Agglomeration," *Journal of Urban Economics* 50 (September): 191–229.

U.S. Bureau of the Census (1996), *USA Counties 1996,* CD-ROM, Washington, D.C.: U.S. Department of Commerce.

__________ (2001), *1997 Economic Census,* CD-ROM, Washington, D.C.: U.S. Department of Commerce.

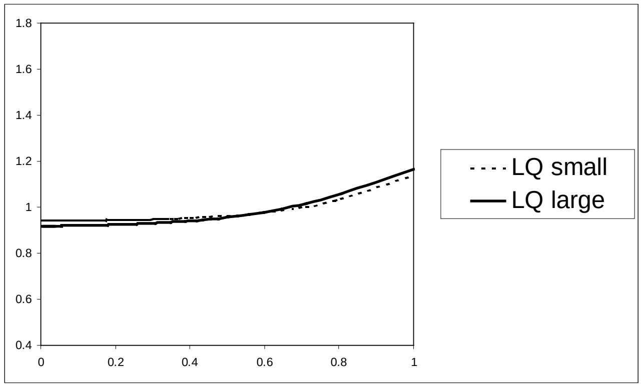
U.S. Bureau of Transportation Statistics (2000), *Public Use Airports,* Washington, D.C.: Department of Transportation.

Figure 1

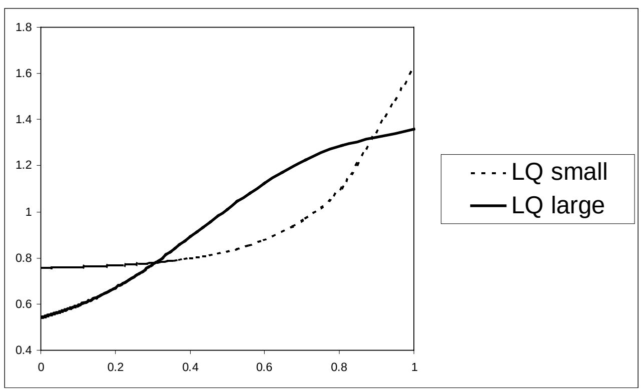
LQ in Model by City Size* for Three Levels of τ

(When $\alpha = 1$ and $J = 100$)

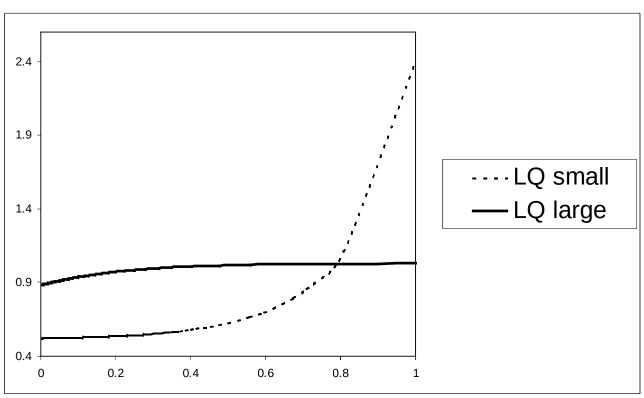
A. **Small τ**



B. **Medium τ**



C. **Large τ**



*In each panel, the *x* variable on the horizontal axis is the cumulative population of cities sorted in ascending population, and the *y* variable is the location quotient of the marginal city when the cumulative population is *x*.

Table 1

Sales-Office Intensity Measures by City Size

		MSA Population (millions)			
	Non-MSA	Under 0.5	0.5–2	2–8	Over 8
Per Capita Measures					
Sales ($1,000 per person)	.62	2.70	4.92	7.98	6.86
Employment (per 1,000 in population)	.99	2.71	3.86	4.86	5.23
Payroll ($1,000 per person)	.03	.11	.18	.27	.28
Operating Expenses ($1,000 per person)	.07	.22	.36	.54	.61
Inventories ($1,000 per person)	.05	.14	.16	.22	.24
Location Quotients					
Sales	.12	.52	.95	1.55	1.33
Employment	.26	.72	.03	1.30	1.40
Payroll	.17	.57	.93	1.44	1.46
Operating Expenses	.19	.57	.92	1.38	1.56
Inventories	.30	.85	.96	1.28	1.40
Number of MSAs	—	194	57	19	3

Source: Author's calculations with publicly available data from the 1997 Census of Wholesale Trade

Table 2

MSA-Level Regressions:
Log Sales on Log Population

	Slope (std. err.)	R^2
Cross-Section Regressions (year)		
1982	1.64 (.05)	.78
1987	1.63 (.06)	.76
1992	1.68 (.06)	.75
1997	1.71 (.06)	.75
1997 With Controls for • Education • Airport Access • Manufacturing Activity	1.56 (.06)	.82
Fixed-Effect Regression (1982–97)	1.80 (.34)	.09

Source: Author's calculations with confidential micro data from the Census of Wholesale Trade, 1982–97

Table 3

Selected Statistics by Sales Size of Firm

Firm Sales Size ($ mil.)	Number of Firms	Number of Offices	Average Offices per Firm	Total Sales of Size Class ($ bil.)	Sales per Office ($ mil.)
Under 25	2,209	4,091	1.9	16.6	4.1
25–49.99	463	1,711	3.7	16.6	9.7
50–99.99	362	2,120	5.9	25.5	12.0
100 and Over	942	21,209	22.5	1,191.0	56.2
All Firms	3,976	29,131	7.3	1,249.7	42.9

Source: Author’s calculations with publicly available data from the 1997 Census of Wholesale Trade

Table 4

Estimated Location Quotients From Logit Model by Sales Size of Firm

		MSA Population (millions)			
Sales of Firm ($ mil.)	Non-MSA	Under 0.5	0.5–2	2–8	Over 8
Under 25	.30	.61	.96	1.31	1.74
25–50	.21	.56	.94	1.38	1.81
50–100	.13	.39	.93	1.58	1.76
100–250	.18	.50	1.08	1.49	1.55
250–1000	.15	.40	1.10	1.55	1.55
Over 1000	.04	.38	1.05	1.80	1.37

Source: Author's calculations with confidential micro data from the 1997 Census of Wholesale Trade

Table 5

Mean Sales Size and Cell Counts by Sales-Size Category

(Includes only offices in MSAs)

Sales of Firm ($ mil.)	Mean Sales ($ mil.)	Number of Firms
Under 25	7.5	2,097
25–50	35.8	426
50–100	70.7	364
100–250	159.1	368
250–1000	479.0	335
Over 1000	4,856.9	196

Source: Author's calculations from confidential micro data from the 1997 Census of Wholesale Trade

Table 6

Structural Parameter Estimates

(1997 Data)

Parameter	*Model 1:* No Additional City Characteristics	*Model 2:* Additional City Characteristics
ϕ	.774 (.014)	1.200 (.032)
τ	1.665 (.017)	1.663 (.020)
γ	1.605 (.231)	2.999 (.355)
$\lambda_{college}$	—	–.0012 (.0003)
$\lambda_{airports}$	—	.027 (.003)
$\lambda_{manufacturing}$	—	.011 (.001)
Number of MSAs	273	273
Number of Firms	3,786	3,786
Chi-Squared Statistic	5,863.95	4,890.56

Source: Author's estimates with confidential micro data from the 1997 Census of Wholesale Trade

Table 7

Comparison of Model 1 With 1997 Census Data

A. **The 1997 Census Data**

Sales of Firm ($ mil.)	Offices per Firm	*Sales Location Quotient* by MSA Population Groupings (millions)			
		Under 0.5	0.5–2	2–8	Over 8
Under 25	1.7	.76	.91	1.04	1.27
25–50	3.3	.62	.78	1.09	1.47
50–100	4.9	.52	.80	1.13	1.48
100–250	7.6	.55	.89	1.16	1.28
250–1000	13.8	.44	.89	1.22	1.29
Over 1000	28.4	.32	.85	1.33	1.28

B. **Model 1**

Sales of Firm ($ mil.)	Offices per Firm	*Sales Location Quotient* by MSA Population Groupings (millions)			
		Under 0.5	0.5–2	2–8	Over 8
Under 25	1.4	.72	.75	.90	1.74
25–50	5.3	.62	.65	.93	1.90
50–100	9.3	.56	.60	1.05	1.81
100–250	15.8	.50	.58	1.17	1.71
250–1000	27.3	.44	.66	1.20	1.61
Over 1000	49.9	.46	.75	1.18	1.52

C. **Model 1 With the γ Parameter Set to Zero**

Sales of Firm ($ mil.)	Offices per Firm	*Sales Location Quotient* by MSA Population Groupings (millions)			
		Under 0.5	0.5–2	2–8	Over 8
Under 25	1.4	.85	.88	.95	1.38
25–50	5.4	.73	.74	.95	1.65
50–100	9.5	.65	.67	1.08	1.60
100–250	16.2	.57	.65	1.18	1.53
250–1000	28.4	.50	.72	1.21	1.45
Over 1000	52.6	.52	.81	1.19	1.38

Table A1

Distribution of City Characteristics by City Size (%)

Characteristic	Non-MSA	*MSA Population* (millions) Under 0.5	0.5–2	2–8	Over 8
Education Level (% of population 25 years and older with 4 or more years of college)	13.26	18.56	20.44	24.64	23.80
Airport Activity (Domestic enplanements per person in 1999)	—	.92	2.76	4.06	2.64
Manufacturing Activity (Sales of manufacturing plants, $1,000 per person in 1999)	14.54	16.65	15.36	14.27	10.86

Table A2

MSA Regression With Population and Other Factors

(Dependent Variable = Log Sales)

Variable	Estimate (std. err.)
Constant	6.50 (.42)
Log Population	1.56 (.06)
Log Education Level	.30 (.21)
Airport Activity	.14 (.03)
Log Manufacturing Activity	.62 (.08)
R^2	.82
N	273

Table A3

Regression Results for Neighboring Population Decile Groupings: Log of Per Capita Measures on Log of Average Neighboring Population

Sales-Office Activity Measure (Per capita)	Slope (std. err.)	R^2
Sales	.64 (.08)	.88
Employment	.45 (.05)	.91
Payroll	.59 (.06)	.92
Operating Expenses	.57 (.05)	.94
Inventory	.44 (.05)	.92