EVOLVING AGGLOMERATION IN THE U.S. AUTO SUPPLIER INDUSTRY*

Thomas Klier

Federal Reserve Bank of Chicago, Research Department, 230 S. LaSalle St., Chicago, IL 60604. E-mail: tklier@frbchi.org

Daniel P. McMillen

Department of Economics (MC 144), University of Illinois at Chicago, 601 S. Morgan St., Chicago, IL 60607. E-mail: mcmillen@uic.edu

ABSTRACT. Using nonparametric descriptive tools developed by Duranton and Overman (2005, *Review of Economic Studies*, 72, 1077–1106), we show that both new and old auto supplier plants are highly concentrated in the eastern United States. Conditional logit models imply that much of this concentration can be explained parametrically by distance from Detroit, proximity to assembly plants, and access to the interstate highway system. New plants are more likely to be located in zip codes that are close to existing supplier plants. However, the degree of clustering observed is still greater than implied by the logit estimates.

1. INTRODUCTION

The North American automobile industry has been remarkably concentrated since its inception. Assembly operations are characterized by significant scale economies in production. Only a small number of assembly plants are required to serve the entire continent, and these plants tend to be located in the center of the country. Although prominent exceptions to this rule were once operating on both the east and west coasts, many of these outlying plants have been closed in recent years as the industry has re-trenched toward the middle of the U.S. and lower Ontario.

These trends, which are documented in Rubenstein (1992) and Klier and McMillen (2006), have been accompanied by changes in the geographic distribution of auto supplier plants. Though supplier plants are often part of comparatively small firms, their operations are also subject to internal scale economies. A supplier plant may serve several assembly plants. Moreover, the rise of just-in-time inventory practices has increased the incentive for suppliers to locate

close to assemblers. Supplier plants thus tend to cluster near assemblers, and suppliers too have re-trenched toward the center of the country in recent years. Maps of assembly and supplier operations show a growing concentration of auto suppliers along an axis running southward from Detroit. Whereas, the industry once was concentrated in a corridor running from Chicago to New York, it now has a north-south orientation.

In this paper, we use both parametric and nonparametric techniques to document the changing geographic structure of the American auto supplier industry. We focus on suppliers rather than assemblers because their much larger number makes them more amenable to statistical analysis. Of the 2,627 supplier plants in our dataset, 431 opened after 1990. Using a nonparametric approach developed by Duranton and Overman (2005), we begin by documenting the degree of localization exhibited by this industry. Both new and old supplier plants are far more concentrated than would be expected by pure randomness, and this result holds whether we define randomness as an equal chance that a plant might locate in any zip code in the eastern U.S. or we weight the probability by the level of employment in the zip code. This descriptive analysis suggests that the geographical distribution of new and old plants is remarkably similar given the amount of change undergone by the auto industry during this time.

The next step in our analysis is a parametric investigation of the determinants of auto supplier locations. We model the location decision of individual plants using a conditional logit model. The availability of plant-level data allows us to utilize the zipcode level of geography. We find that both new and old plants are more likely to be located in zip codes that are near assembly plants, close to Detroit, and are served by interstate highways. In addition, we find that new plants are more likely to be located in zip codes that are close to existing supplier plant locations. The changing geographic orientation of the industry is evident in the conditional logit estimates: new plants are more likely than existing plants to locate in the East South Central region. However, the similarities are more striking than the differences. As new plants open in the southern United States, they tend to follow a location pattern similar to the plants that have preceded them in the region.

In the final step of our analysis, we use the predicted probabilities from the conditional logit models as the base for the Duranton and Overman (2005) measure of concentration. We find that actual plant locations are even more concentrated than implied by the conditional logit estimates. However, new plant locations are *not* more concentrated than would be implied by a simple random choice from existing plant locations. This result reinforces our finding that new plants follow a location pattern similar to existing plants. We also find that Duranton and Overman's (2005) nonparametric procedure is useful as a diagnostic tool: the conditional logit models, while apparently fitting the data well, fail to account adequately for the degree of clustering exhibited in practice.

Whether the focus is on new or old plants, our results portray a highly clustered auto supply industry. Plants opening after 1990 are more likely than older plants to locate along an axis running south from Detroit. But both new

and old plants are highly concentrated, locating close to assembly plants, near highways, and near other supplier plants. Although the geographic orientation has moved south, the industry is neither more nor less concentrated now than prior to 1990.

2. THE U.S. AUTO INDUSTRY

In the 1890s, during the beginnings of the U.S. auto industry, more than half of the producers of automobiles were located in the northeast between Philadelphia and Boston.[1] Soon afterwards, during the first decade of the twentieth century, southeastern Michigan emerged as the hub of auto production in the U.S. It attracted or retained the most successful motor vehicle producers because many of the industries from which automotive technology is derived, such as the production of engines and carriages, were already thriving in the region.[2] Subsequently, automakers and suppliers could tap into a rich pool of skilled mechanics and engineers. According to the 1904 Census of manufacturers, 42 percent of all cars were made in Michigan, as the industry's leading producers and their major facilities were based in Michigan by then.

Over time, the location of auto assembly and auto parts plants evolved differently. The Ford Motor Company developed a system of branch assembly plants which was quickly copied by the other major producers of vehicles. It was based on the fact that auto assembly is a classic weight-gaining industry: it is cheaper to produce finished vehicles near the centers of population than to ship finished vehicles from a central location to many destinations across the country. Motor vehicle parts, on the other hand, continued to be produced in the Midwest and then shipped to the various assembly plants located across the country. A quickly growing industry was well-suited for a branch assembly plant system as production runs for the best-selling vehicles were large enough to support more than one assembly plant. This combination of decentralized assembly plants combined with the concentration of parts production in Michigan and its neighboring states of Indiana and Ohio continued until the 1980s.

The forces leading to a re-structuring of the auto industry geography began during the 1960s. In response to increased sales of smaller cars by foreign producers, the U.S. producers introduced a number of smaller platforms over the years, for example, "subcompact," "compact," and "intermediate" cars. As a result the growth of product variety outpaced the growth of overall demand, leading to substantially smaller production runs, even for the best-selling models. Subsequently, no individual model sold enough to justify production at

---

[1]This section draws heavily on Rubenstein (1992).

[2]Bicycle manufacturing, the third major contributor to the early development of the automobile, proved the exception as the country's largest bicycle manufacturers were located in the Northeast. According to Rubenstein (1992), bicycle manufacturers contributed to the emergence of southeastern Michigan as the industry's hub by failing to recognize the automobile's potential and thereby losing their early technological lead in the face of rapid technological innovation.

more than one, or at most two, assembly facilities. This development led to a re-concentration in the geography of auto production. In conjunction with the recessions induced by the 1970s oil crises and an increase in motor vehicle imports, domestic auto producers reduced capacity and shut down some of their production facilities. Specifically, assembly plants located on the coasts were increasingly abandoned in favor of locations in the center of the country. As a result, the location of assembly plants began to re-concentrate in the Midwest.

Starting in the early 1980s, foreign producers began producing vehicles in the United States.[3] They strongly preferred locations in the interior of the country. Yet the foreign producers extended the auto region to the south by opening plants in Kentucky and Tennessee, and most recently as far south as Mississippi and Alabama (see Klier and McMillen, 2006).

The auto industry has experienced a significant southward extension even as the Midwest re-emerged as the center of vehicle production after the demise of the branch plant system in vehicle assembly. This movement southward has been driven primarily by the location of foreign-owned assembly plants during the 1980s and 1990s. Incidentally, most of these plants are located at greenfield sites, some distance from traditional manufacturing locations. Today, the preferred locations for motor vehicle assembly are defined by a north-south region that is often referred to as the I-65–I-75 corridor, as it is rather well defined by two of the major north-south interstate highways, extending south from Michigan to Tennessee and beyond.

Thus, North American auto supplier plants have been remarkably concentrated for a long time (Klier and McMillen, 2006). When the industry got its start just over 100 years ago, raw materials and worker skills available in the upper Midwest, between Chicago and Buffalo, furthered the development of this industry. Auto suppliers remained concentrated in the upper Midwest during the branch (assembly) plant era, as it was cheaper to ship parts than finished vehicles from a central location. During the early 1980s the U.S. auto industry was shaped by the arrival of foreign producers who brought with them the Just-in-time production system as well as a substantial number of foreign suppliers. The 1980s also witnessed the emergence of the auto corridor, a region extending south from Detroit into Kentucky and Tennessee, with fingers reaching into Mexico and Canada. During this time new parts plants showed a tendency to locate farther south, reinforcing the north-south orientation of the auto region.

This brief overview of the geography of the U.S. auto industry shows a long-clustered industry that now remains highly clustered after a recent major re-orientation southward.[4] The spatial concentration of today's industry

---

[3]The exception is Volkswagen, which started producing cars in Westmoreland, PA, in 1978. The company's spell of producing cars in the U.S. did not last very long. That plant closed in 1989.

[4]Klier and McMillen (2006) trace in some detail the re-orientation of the auto industry geography by comparing location choices for assembly and supplier plants during the 1980s and 1990s. They also compare the location patterns of domestic and foreign plants. Woodward (1992)
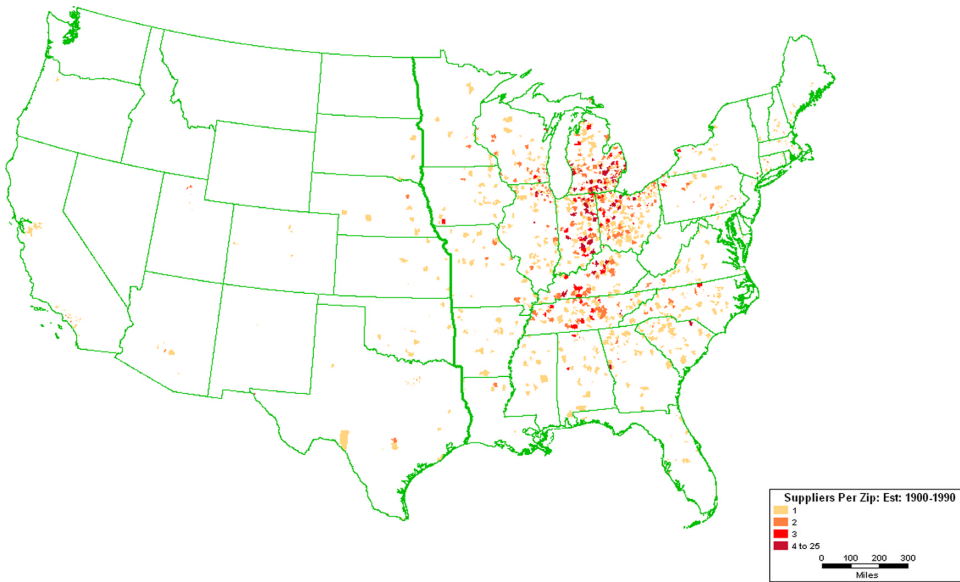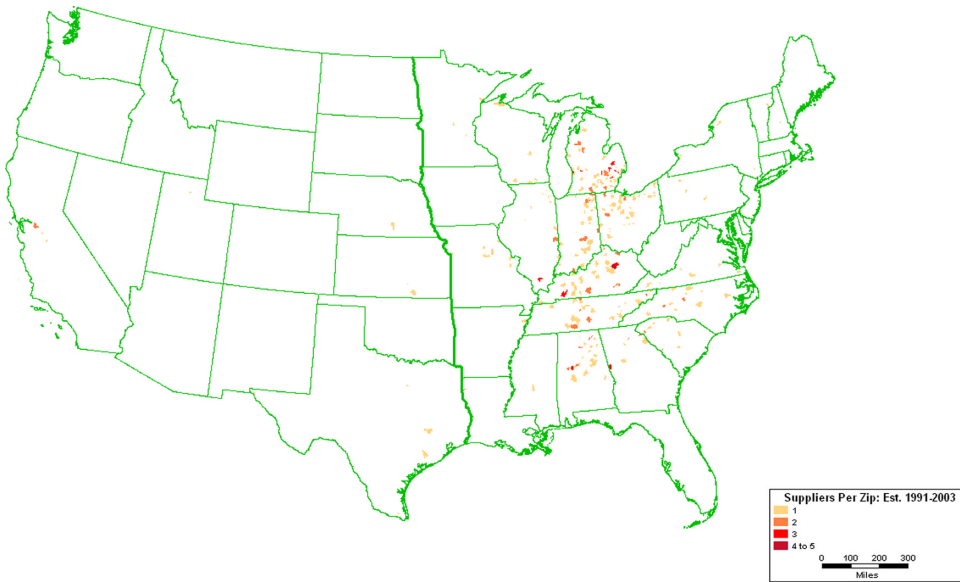
Source: ELM International, state manufacturing directories, supplier company websites, Harbour Consulting (2003), and Maptitude. **This figure appears in color in the online version of the article [DOI: 10.1111/j.1467-9787.2007.00549.x]**

FIGURE 1: Distribution of Old Auto Supplier Plants.

(Ellison and Glaeser, 1997) is reinforced by tightly linked supply chains that require most suppliers to be within a day's shipping distance of their assembly plant customers. Figures 1 and 2 illustrate the changing geography of auto supplier plants. Both maps are based on the data used later in our statistical models. The maps show an industry concentrated in the eastern half of the U.S., where the vast majority of plants producing auto parts destined for vehicle assembly are located. Figure 1 shows the distribution of "old" auto supplier plants. The most densely populated zip codes define a north-south auto region, with southern Michigan, Indiana, and Ohio as its hub. Yet the industry covers a much larger area as its plants are well represented in almost every state in the eastern part of the United States. Though Figure 2 is based on a much smaller number of plants (1/5 of the number of plants represented in Figure 1), it clearly illustrates the formation of a rather well-defined auto region that extends south from Michigan to northern Alabama and Georgia, reaching into the Carolinas. These maps clearly show that auto supplier plants that opened between 1991 and 2003 re-trenched toward the center of the country.

---

and Smith and Florida (1994) find evidence that vertical linkages as well as the presence of highway infrastructure influence plant location decision of Japanese plants in the United States.

Source: ELM International, state manufacturing directories, supplier company websites, Harbour Consulting (2003), and Maptitude. **This figure appears in color in the online version of the article [DOI: 10.1111/j.1467-9787.2007.00549.x]**

FIGURE 2: Distribution of New Auto Supplier Plants.

3. DATA

A Michigan-based vendor, ELM International, provided the primary data for our analysis. The ELM database, which covers the entire North American auto industry, includes suppliers that ship directly to assembly plants as well as suppliers that ship indirectly (via other suppliers) to assembly plants. Our analysis pursues only plants located in the eastern United States. As Figures 1 and 2 illustrate, the geography of the 48 contiguous states is substantially larger than the geography of the U.S. auto supplier industry. In fact, the auto industry is absent from large expanses of the western United States. Whereas, the eastern Unites States can be treated conceptually as an integrated market area, the western states appear to be quite distinct. Therefore, we limit our analysis to the eastern United States. While we include states that border the western bank of the Mississippi River in order to include large concentrations of plants in places such as St. Louis we omit states to the west which include very few plants.[5] The 31 states represented in our definition of the eastern United States form a reasonably compact and integrated economic area.

The ELM database includes data at the plant and company level. However, plants that produce machine tools or raw materials and those that produce

---

[5]The boundary of our area of study is indicated in Figures 1 and 2 by the accentuated line of state borders running north from Texas' eastern border.

primarily for the aftermarket are not part of the database. The data include information on "captive" supplier plants, which are parts operations that assemblers own and operate themselves, such as engine and stamping facilities. The database includes information on a plant's address, products, employment, parts produced, customer(s), union status, as well as square footage. Records were cross-checked with state manufacturing directories to obtain information on the plant's age, and information on captive plants was obtained from Harbour (2003). We then geocoded the data to the zip code level and verified the accuracy of the data whenever possible by checking individual company's websites and through phone calls.

The dataset includes data for 3,298 supplier plants in the eastern United States.[6] Of these plants, 431 are "new," which we define as having opened since 1991. For 671 observations we could not obtain data on plant age. The relevant characteristics of the locations of these plants differ from both the locations of existing and new plants (see Table 1). Since it is unclear whether these plants are old or new, we adopted a conservative strategy of dropping them from our analysis altogether. We refer to the remaining 2,196 observations, which began operation before 1991, as "existing" or "old" plants. Since the dataset is cross-sectional in nature, the age variable applies only to surviving establishments. Although this focus on survivors may lead us to understate the extent to which "old" plants are concentrated near Detroit, it provides an accurate view of the geographic distribution of new plants and it allows us to test whether the distribution differs from that of surviving older plants.

Using 1991 as the starting date for new plants allows us to determine whether the major changes undergone by the American auto industry in the 1980s fundamentally altered the geographic distribution of the industry. A further advantage of focusing on plant openings from after 1990 is that it allows us to match the plant openings with explanatory variables from the 1990 U.S. Census. Moving the date forward by 1 year from the time of the census ensures that these explanatory variables can be taken as exogenous.

Table 1 presents descriptive statistics for the variables used in our analysis. Separate sets of statistics are presented for the new and old plant samples. In addition, we present descriptive statistics for samples of randomly chosen alternative locations. These alternative locations comprise the rejected alternatives for our conditional logit models. To identify these alternatives, we match each plant with five randomly chosen zip codes that (1) are different from the plant's actual zip code, and (2) are different from each other. Candidate alternatives include any zip code in the eastern United States, including those with neither a new nor old supplier plant. The alternative for one plant may include a zip code that already has another plant.

---

[6]In comparison, the 2002 version of County Business Patterns reports 4,000 plants in the motor vehicle parts sector (NAICS 3,363), which suggests that our database has very good coverage of this industry.

TABLE 1: Descriptive Statistics

	Existing plants		New plants		Age unknown
	Plant zip code	Random alternative	Plant zip code	Random alternative	Plant zip code
Interstate highway	0.5168	0.3186	0.6125	0.3197	0.6215
	(0.4998)	(0.4659)	(0.4877)	(0.4665)	(0.4854)
Distance to nearest assembler (100 miles)	0.6316	1.1662	0.5958	1.1587	0.8168
	(0.6550)	(0.9962)	(0.5566)	(1.0176)	(0.7775)
Number of assemblers within 100 miles	7.0724	1.9405	7.0325	2.1276	6.614
	(8.4403)	(3.6457)	(8.6411)	(4.0782)	(9.0696)
Population density (1,000s per sq. mile)	1.4534	1.6382	1.1125	1.6635	1.3248
	(2.6110)	(7.2710)	(2.1005)	(7.4259)	(2.0537)
Proportion white	0.9027	0.8773	0.8941	0.8799	0.8504
	(0.1741)	(0.2035)	(0.1758)	(0.2029)	(2.0537)
Proportion high school graduates	0.7355	0.7126	0.7200	0.7144	0.724
	(0.1233)	(0.1347)	(0.1238)	(0.1344)	(0.1308)
Proportion manufacturing	0.2713	0.2009	0.2684	0.2006	0.2601
	(0.0932)	(0.1075)	(0.0880)	(0.1070)	(0.0976)
Metropolitan	0.6699	0.5932	0.6845	0.5810	0.7437
New England	0.0291	0.0813	0.0093	0.0701	0.0417
Middle Atlantic	0.0515	0.1802	0.0302	0.1712	0.0537
West North Central	0.0360	0.1117	0.0302	0.1077	0.0432
South Atlantic	0.1015	0.2586	0.1276	0.2488	0.2131
East South Central	0.1120	0.1138	0.2483	0.1118	0.1744
West South Central	0.0109	0.0479	0.0070	0.0501	0.0179
Right to Work state	0.1890	0.3557	0.2645	0.3480	0.3741
Distance from Detroit (100 miles)	2.6959	4.7679	2.8686	4.6955	3.4804
	(2.0829)	(2.2097)	(2.0634)	(2.2894)	(2.4648)
Number of observations	2196	10980	431	2155	671

Note: Standard deviations are in parentheses for the continuous variables.

Table 1 includes descriptive statistics for explanatory variables for the conditional logit models. Having an interstate highway run through a zip code increases the likelihood of having a plant, and the effect is stronger for new plants. Zip codes with either new or old plants are more likely than randomly chosen alternatives to be near assembler plants. Variables drawn from the 1990 U.S. Census include population density, the proportion of the zip code's white population, the proportion who have graduated from high school, and the proportion who work in manufacturing jobs. We also include regional dummy variables and a variable indicating whether the zip code is located in a metropolitan area. Finally, we include a variable measuring the distance in from Detroit. Plants are much more likely to be located close to Detroit and in the base region, the East North Central region.

4. THE GEOGRAPHIC DISTRIBUTION OF SUPPLIER PLANTS

In this section, we use the methodology developed by Duranton and Overman (2005) to compare the geographic concentration of existing and new auto supplier plants in the eastern United States. Our dataset is geocoded down to the zip code level. Using the geographic coordinates, we begin by calculating the distance between every pair of plants. With n plants, there are $n(n-1)$ distance pairs. Using a standard kernel density function (Silverman, 1986), we can calculate the density of bilateral distances at any target distance d as:

$$(1) \qquad K(d) = \frac{2}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{d - d_{i,j}}{h}\right)$$

where $d_{i,j}$ is the distance between observations i and j, h is the bandwidth, and f is the kernel function. As in Duranton and Overman (2006), we use a standard Gaussian kernel with an optimal bandwidth.[7] All distances are measured in straight-line miles. Following Duranton and Overman, we refer to the estimated functions as *K-densities*.

As written, equation (1) will produce positive density estimates at negative distances, which means that the estimated function does not integrate to one over the relevant set of distances. Following Duranton and Overman (2005), we use Silverman's (1986) reflection method to impose zero densities at negative densities. This method involves supplementing each observation with its negative value, $-d_{i,j}$, to form a pseudo data with twice the original number of observations. The revised formula for the K-density estimate simplifies to:

---

[7]To calculate the optimal bandwidth, we first calculate the standard deviation (s) of the $n(n-1)$ bilateral distances. Following Silverman (1986), the optimal bandwidth for a Gaussian kernel is $1.06sn^{-.2}$.

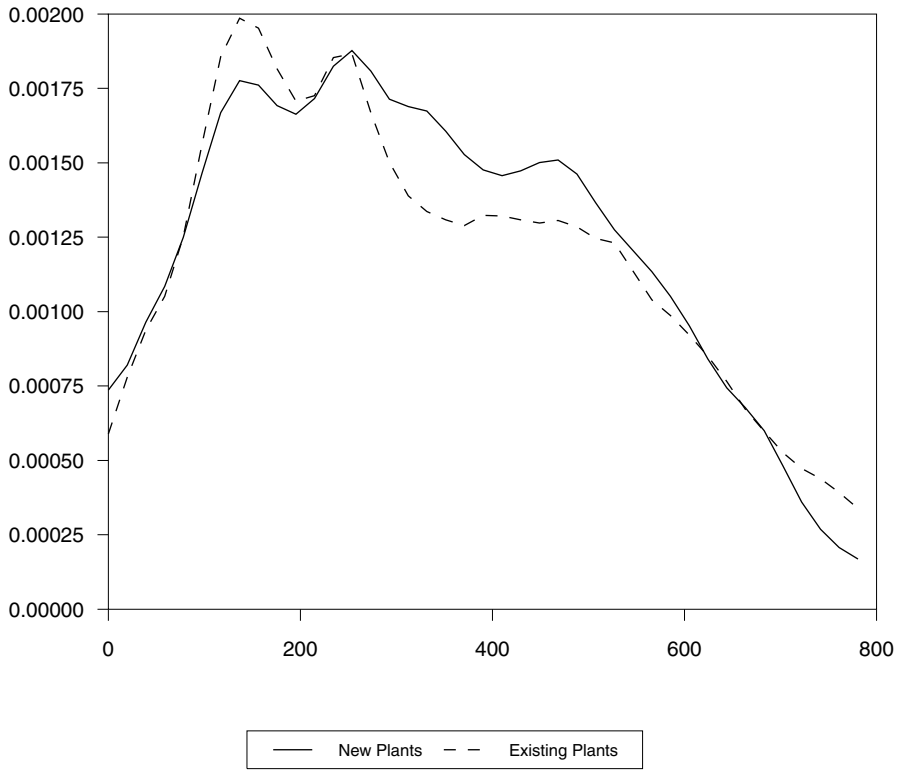FIGURE 3: Densities for Existing and New Supplier Plants.

$$(2) \quad K(d) = \frac{2}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[f\left(\frac{d - d_{i,j}}{h} \right) + f\left(\frac{-d - d_{i,j}}{h} \right) \right] \quad \text{for } d > 0$$

$$K(d) = 0 \qquad\qquad\qquad\qquad \text{for } d \leq 0.$$

This formula ensures that the estimated densities integrate to one by producing higher estimates at low distances.

We calculate separate K-densities for new and existing plants. We calculate equation (2) at 41 evenly spaced target points between $d = 0$ and $d = 800$. The results are shown in Figure 3. The striking feature of Figure 3 is the similarity between the estimated densities. Both density functions have twin peaks at distances of about 135 miles and 250 miles. The densities rise rapidly to the first peak and trail off slowly at distances beyond 250 miles. The most common distances between plants are in the range of about 100–300 miles. Given the size of the eastern United States, these distances are not small. Most importantly, the distribution of distances between plants has not changed significantly since 1991. Plants are not substantially closer to one another now than they were before 1991.

Although Figure 3 shows that the K-densities are similar for new and old plants, it does not show directly whether the auto supplier industry is heavily concentrated. Measuring geographic concentration requires a base model of possible locations. To measure concentration, Duranton and Overman (2006) compare actual K-densities to the density that would be expected if plants were located randomly across space. Using a different but related approach, Ellison and Glaeser (1997) compare actual locations to the locations expected if plants were assigned to locations based on the an area's share of total manufacturing employment.

In this section, we use three base models of possible locations to measure geographic concentration. In the first model, the probability that a zip code is chosen as a plant location is $p_i = 1/n_z$, where n_z is the number of zip codes. In the second model, the probability for zip code i is $p_i = E_i/(\Sigma_i E_i)$, where E_i represents total employment in zip code i. Analogously to Ellison and Glaeser (1997) and Duranton and Overman (2005), the probabilities in the third model are based on the share of total manufacturing employment, $p_i = E_{mi}/(\Sigma_i E_{mi})$, where E_{mi} represents zip code i's manufacturing employment. Of the 28,036 zip codes in the eastern United States, 19,506 have some employment while 19,151 have some manufacturing employment.[8]

After assigning a probability of p_i to each zip code, we make n draws without replacement from the set of zip codes to construct our base model set of locations. We then calculate the distance between every actual plant location and the randomly drawn set of replacement locations. We then re-calculate the K-density as:

$$(3) \qquad K(d) = \frac{1}{n^2 h} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[f\left(\frac{d - d_{i,j}}{h} \right) + f\left(\frac{-d - d_{i,j}}{h} \right) \right]$$

where $d_{i,j}$ denotes the distance between the actual plant location i and the randomly drawn location j. There are now n^2 distances to calculate—n base plant locations and n zip code draws. However, we maintain h at the value used in equation (2) to keep the level of smoothing at the same level as before. We use this bootstrap re-sampling procedure to generate 2,000 new estimates of the K-density functions. Each estimated function is estimated at each of the 41 target distances.

Let $\hat{K}_b(d)$ represent the estimated K-density at target distance d for bootstrap sample b, where $b = 1, \ldots, B$ and $B = 2,000$. Standard 95 percent bootstrap confidence intervals can be calculated by ordering the values of $\hat{K}_b(d)$ from smallest to largest and choosing the 50th and 1950th largest values. Duranton and Overman (2005) refer to the result of this calculation as a *local* confidence interval. At each target distance d, 95 percent of the estimated K-densities fall

---

[8]Our approach differs slightly from Duranton and Overman (2005), who base their counterfactual density on the presence of manufacturing firms rather than the number of manufacturing employees. Our approach implies that zip codes with large firms are more likely to attract new firms than sites with low levels of employment.

within the interval. However, the full K-density functions have 41 values, one for each target distance. A function lying within the local confidence interval at one target distance may well fall outside the local interval at another distance. Duranton and Overman (2005) propose an alternative procedure to construct a *global* confidence interval, which takes into account the fact that the each function is estimated simultaneously at multiple target points. The objective of the global confidence interval is to find a band such that no more than 95 percent of the estimated density functions have even a single value that lies outside the interval at any of the target distances.[9]

The construction of the *local* confidence intervals can be expressed mathematically as follows. At each target location d, the lower bound for the 95 percent confidence interval is the value $l_1(d)$ such that $\sum_{b=1}^{B} I(\hat{K}_b(d) < l_1(d)) \leq 0.025B$, where $I(\bullet)$ is a function indicating that the condition is true. The upper bound of the 95 percent local confidence interval at location d is the value $h_1(d)$ such that $\sum_{b=1}^{B} I(\hat{K}_b(d) > h_1(d)) \leq 0.025B$. In contrast, the *global* confidence intervals, $[l_2(d), h_2(d)]$, treat each of the B estimated density functions as single observations. If the estimated density function for bootstrap draw b falls below the global lower bound at any of the target distances, then $\max I(\hat{K}_b(d) < l_2(d)) = 1$. The lower bound for the global 95 percent confidence interval is defined as a set of D values for $l_2(d)$ such that $\sum_{b=1}^{B} \max I(\hat{K}_b(d) < l_2(d)) \leq 0.025B$, where D is the number of target distances ($D = 41$ in our application). Similarly, the upper bound for the global confidence interval is a set of D values $h_2(d)$ such that $\sum_{b=1}^{B} \max I(\hat{K}_b(d) > h_2(d)) \leq 0.025B$.

The global confidence intervals are clearly larger than the local intervals. Whereas the local intervals are found by sorting the values of $\hat{K}_b(d)$ and choosing the $0.025B$ and $0.975B$ largest values, the search for the global bands moves outward from these critical points. After ordering the $\hat{K}_b(d)$ from smallest to largest, define $\psi_c(d)$ as entry c. For the lower bound, we begin by calculating $\sum_{b=1}^{B} \max I(\hat{K}_b(d) < \psi_{50}(d))$. If this value exceeds $0.025B$, we move to entry $c = 49$ (i.e., $0.025B - 1$), and test whether $\sum_{b=1}^{B} \max I(\hat{K}_b(d) < \psi_{49}(d))$ exceeds $0.025B$. Thus, the critical points are being lowered congruently for each of the D target distances. The critical points continue to be lowered until we find the value of c, denoted by c^*, such that $\sum_{b=1}^{B} \max I(\hat{K}_b(d) < \psi_{c}^{*}(d)) \leq 0.025B$. The lower bound at each d is simply entry c^* of the ordered vector of $\hat{K}_b(d)$, i.e., $l_2(d) = \psi_{c}^{*}(d)$. Analogous calculations are made for the upper bound to ensure that $\sum_{b=1}^{B} \max I(\hat{K}_b(d) > \psi_c(d)) \leq 0.025B$: starting at entry $0.975B$ of the ordered vector of $\hat{K}_b(d)$, raise the critical point to entry $0.975B + 1$, $0975B + 2$, and so on until we find the critical value that makes the inequality hold.[10]

---

[9]We are indebted to Gilles Duranton and Henry Overman for their patient explanation of their procedure for calculating global confidence intervals.

[10]In our application, the critical values hover near entries 3 and 1998 of the ordered bootstrap K-density values. Although the critical values come close to the limits of 1 and 2,000, the resulting global confidence intervals are quite close to the local intervals.
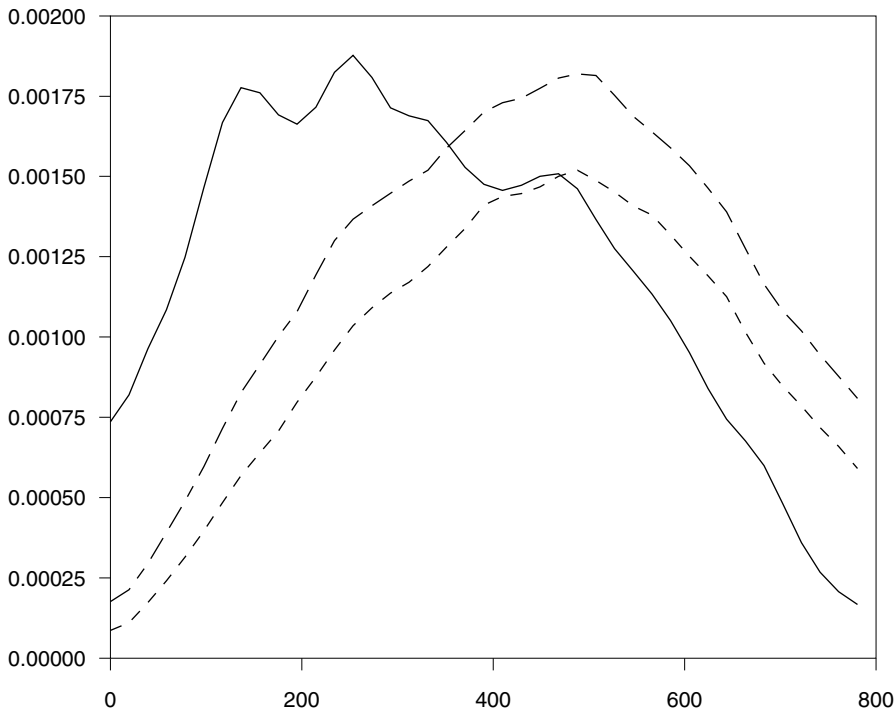
FIGURE 4: New Supplier Plant Density and Confidence Interval-
Random Choice of Any Zip Code.

Figure 4 shows the actual new supplier plant K-density (the solid line) and the 95 percent bootstrap global confidence interval for the K-density based on the uniform probabilities $p_i = 1/n_z$. Given the large area covered by the eastern United States, simply assigning 431 new plants randomly across space would lead to a density function with a peak at a distance of roughly 500 miles. Instead, the actual K-density function has twin peaks at about 135 miles and 250 miles. The actual K-density function is well above the 95 percent global confidence interval from distances of zero to 350 miles. Figure 4 provides clear visual evidence that new supplier plants are highly concentrated geographically.

Figures 5 and 6 show comparable confidence intervals for the new-plant K-density function based on the total employment probabilities $E_i/(\Sigma_i E_i)$ and the manufacturing employment probabilities $E_{mi}/(\Sigma_i E_{mi})$. The 95 percent global confidence intervals are virtually identical because the two sets of probabilities are highly correlated.[11] The only difference between these figures and Figure 4

---

[11]Across all 28,036 zip codes, the correlation between the total and manufacturing employment shares is 0.87. The correlation is 0.86 for the 19,151 zip codes that have some manufacturing employment.
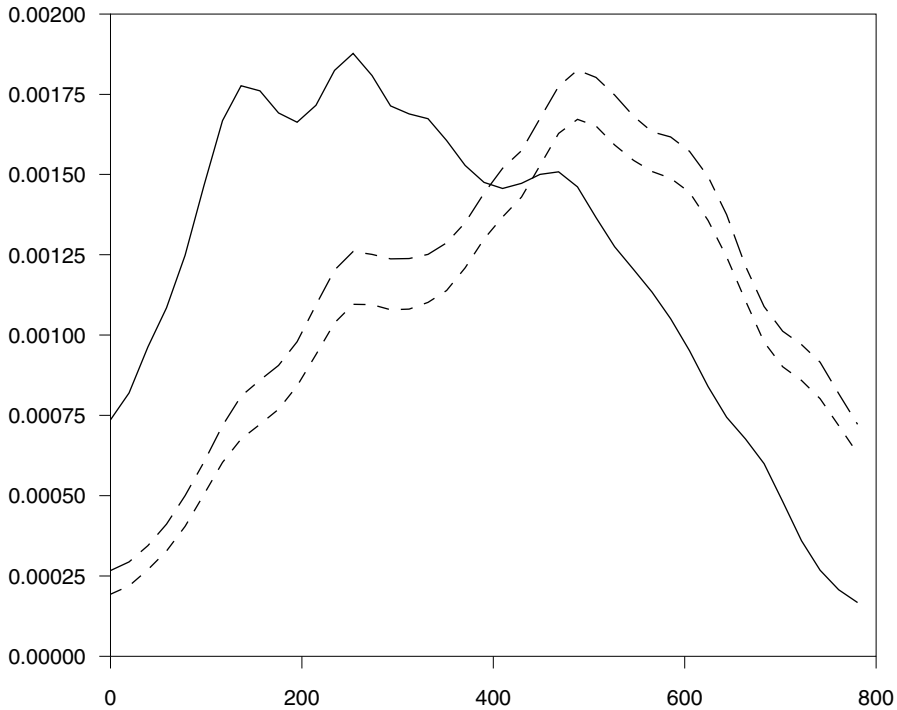
FIGURE 5: New Supplier Plant Density and Confidence Interval-
Total Employment.

is that the area where the actual K-density function is above the 95 percent confidence interval extends a bit farther—to 390 miles rather than 350. Whether we use uniform probabilities or weight the probabilities by employment shares, the K-densities imply a highly concentrated distribution of auto supplier plants in the eastern United States.

5. CONDITIONAL LOGIT MODELS OF PLANT LOCATIONS

In this section, we present conditional logit models explaining the probability that an auto supplier plant is located in a zip code. The primary question is whether we can explain the geographic concentration of supplier plants with such key explanatory variables as distance from Detroit, the presence of a highway, and proximity to assembly plants. Our analysis is not the first attempt to model the location decision of auto supplier plants. Woodward (1992) and Smith and Florida (1992) use county-level data to establish the importance of highway transportation as a determinant of plant location. However, our analysis is unique in the level of geographic detail and the use of a conditional logit approach in place of a simple multinomial logit. With 28,036 zip codes, 2,627 plants, and plant openings as recent as 2003, our dataset is unusually detailed.
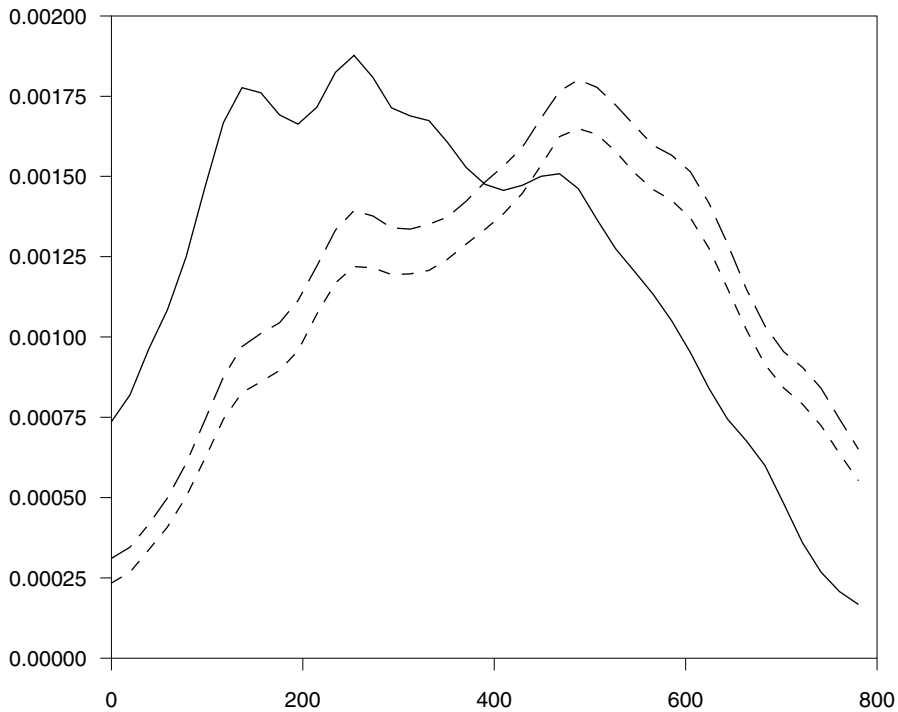
FIGURE 6: New Supplier Plant Density and Confidence Interval-
Manufacturing Employment.

The existing literature uses county-level data and multinomial logit models to determine the effect of county characteristics on the probability of plant location. In contrast, the conditional logit model operates at a more micro level. For each plant, we know the characteristics of rejected zip codes as well as the characteristics of the chosen location. The conditional logit model combines the chosen and rejected locations to produce a much more efficient set of coefficient estimates. Implicitly, each plant faces 28,036 potential location choices. However, we follow Ben-Akiva and Lerman (1985) and randomly choose five rejected alternatives when estimating the model. Since the rejected alternatives are chosen randomly, the resulting coefficients estimates are consistent and more efficient than a simple county-level multinomial logit model.[12]

---

[12]The alternative would be to estimate a simple binomial logit model in which the dependent variable equals one if a plant is located in a county. The assumption behind the standard binomial logit model is that the presence of plant in one county is independent of any information from other counties. Since plants are implicitly choosing among counties, adding information from rejected counties allows the conditional logit model to more accurately identify a variable's effect on the decision process. In practice, the marginal-effect estimates from conditional and standard logit

The resulting dataset has $6n$ observations. The dependent variable equals 1 for the first observation for each plant and the explanatory variables include the characteristics for the chosen plant location. The dependent variable equals zero for the next five observations for each plant and the explanatory variables include the characteristics for the randomly chosen rejected locations. The standard errors are adjusted for the clustering that is implicit in having six observations for each plant.

The results are shown in Table 2. For existing plants—those that opened prior to 1991—the results imply that a zip code is more likely to be chosen as a plant location if an interstate highway runs through it,[13] assemblers are nearby, it is in a metropolitan area, and it is in a right to work state.[14] The probability of an existing plant is higher when the zip code is near Detroit, has a high-population density and a high proportion of manufacturing in employment. All of the regional dummy variables are significant except East South Central, whose negative coefficient is not significantly different from zero.

In the second column of results in Table 2, the specification for new plant locations is similar to the model of existing plant locations. However, we add as an explanatory variable the number of existing supplier plants within 100 miles. This variable can reasonably be taken as exogenous for the new-plant model. Unfortunately, the number of assemblers and the number of suppliers are very highly correlated: the correlation between these two variables is 0.91 in the zip codes with new plants. This multicollinearity makes it difficult to separate the effects of proximity to assemblers and to existing supplier plants. In the last column of results, we present the results when the model is re-estimated after keeping only the more influential variable, the number of existing supplier plants. Making allowances for the smaller sample size, the results for new plants are very similar to the results for existing plants. A zip code is more likely to have a new plant if it is served by an interstate highway, is close to an assembler, has a high proportion of employment in manufacturing, and is close to Detroit. We also find that new plants are more likely to choose zip codes that are within 100 miles of existing industry plants, both models also suggests that, for new supplier plant locations, proximity to the nearest assembly plant matters instead of the number of assembly plants that are within 100 miles. This variation in the way existing assembly plant locations affect the choices of supplier plants is consistent with evidence of tighter linkages between assemblers and suppliers during the 1990s. An increasing number of logistics and

---

models do not differ greatly, although the additional precision of the conditional model results in lower standard errors. Though the results do not vary substantially, we prefer the conditional logit approach because it more closely matches the firm's choice problem in asking why a county was chosen over a set of alternative locations.

[13]Conceptually, the presence of transportation infrastructure supports linkages between assemblers and suppliers as well as among suppliers.

[14]In states with right to work laws, a worker does not have to join a union as a condition for working in a unionized plant. Since nearly all right to work states are in the South, it is sometimes difficult to distinguish the effects of this variable from regional indicators.

TABLE 2: Conditional Logit Models

	Existing plants	New plants	New plants
Interstate highway	0.6957*	1.2154*	1.2154*
	(0.0609)	(0.1450)	(0.1450)
Distance to nearest assembler (100 miles)	−0.0114	−0.4203*	−0.4126*
	(0.0604)	(0.1470)	(0.1435)
Number of assemblers within 100 miles	0.0340*	−0.0068	
	(0.0060)	(0.0281)	
Population density (1,000s per sq. mile)	0.0115*	−0.0037	−0.0037
	(0.0051)	(0.0219)	(0.0220)
Proportion white	−0.2627	0.2647	0.2625
	(0.1815)	(0.4076)	(0.4078)
Proportion high school graduates	1.6900*	0.9358	0.9576
	(0.2928)	(0.6804)	(0.6749)
Proportion manufacturing	5.4731*	5.7246*	5.7663*
	(0.3194)	(0.7976)	(0.7790)
Metropolitan	0.2154*	0.3009	0.2996
	(0.0681)	(0.1611)	(0.1610)
New England	−0.7605*	−0.6020	−0.6151
	(0.1739)	(0.5987)	(0.5960)
Middle Atlantic	−1.5705*	−1.2945*	−1.3213*
	(0.1160)	(0.3577)	(0.3400)
West North Central	−0.7301*	0.0557	0.0326
	(0.1552)	(0.4063)	(0.3944)
South Atlantic	−0.9349*	0.4681	0.4489
	(0.1337)	(0.3053)	(0.2943)
East South Central	−0.1737	1.4789*	1.4686*
	(0.1363)	(0.2755)	(0.2719)
West South Central	−0.6324*	−0.0625	−0.0759
	(0.2714)	(0.7443)	(0.7415)
Right to Work state	0.4255*	0.2678	0.2712
	(0.1249)	(0.2801)	(0.2796)
Distance from Detroit (100 miles)	−0.2782*	−0.2770*	−0.2818*
	(0.0309)	(0.0891)	(0.0868)
Number of existing supplier		0.0022	0.0019*
plants within 100 miles		(0.0014)	(0.0007)
Pseudo-R^2	0.3532	0.4076	0.4076
Number of observations	13,176	2,586	2,586

Notes: Standard errors are in parentheses. An asterisk indicates significance at the 5% level.

supplier functions must be performed in very close proximity to the assembly location. In a number of cases, this tendency has led to the construction of a supplier park immediately adjacent to an assembly plant.

   With pseudo-R^2s in the range of 0.35–0.41, the models fit the data well by the standards of discrete choice models. The models suggest the roots of geographic concentration lie in (1) highway access, (2) the desire to locate near assembly plants, and (3) the strong influence of Detroit on location decisions

in the auto industry. In addition, we find that existing supplier plants appear to have some influence on the location of new plants. However, the models are not able to determine whether existing plants exert a causal influence on new plants due to direct agglomerative forces or if existing plant locations are serving as a proxy for missing variables that influence both new and existing plant location choices.

6. K-DENSITIES BASED ON LOGIT PROBABILITY ESTIMATES

The estimated probability estimates from the conditional logit models can be used as the basis for K-density confidence intervals. The implied base model asks a different question than before: are actual plant locations more concentrated geographically than implied by the estimated logit models? If we base the analysis of new-plant K-densities on the new-plant logit probabilities, we have what amounts to a specification test of our conditional logit model. If the model adequately accounts for the determinants of new plant locations, then the actual K-densities for new plants should lie within the 95 percent confidence interval implied by the estimated probabilities. The question is somewhat different, if we base the confidence intervals for new-plant K-densities on the estimated probabilities produced by the logit model of *existing* plant locations. In this case, the question is whether new plants effectively follow the decision rule that is implied by the existing plant model. New plants may seek out locations that have a high probability of having an existing plant even if a plant has not yet located there. In this case, the K-densities for new plants may lie within the 95 percent confidence interval implied by the existing plant logit model even if it differs from the distribution of actual existing plant distance densities.

To calculate the 95 percent global confidence interval for the new-plant K-densities based on the estimated logit models, we again draw randomly without replacement from the set of actual zip codes. The probabilities are based on the estimated conditional logit models. Thus, equation (3) again forms the basis for the bootstrap K-density. Unlike binomial logit, the conditional logit model does not produce an intercept. Instead, separate intercepts are implied for each plant.[15] Many zip codes are not represented in either the set of actual plant locations or the randomly drawn alternatives. To construct probabilities for every zip code in the eastern United States, we take the sample of $6n$ observations and re-estimate the model using simple logit. In the re-estimated model, the dependent variable equals one for n observations and zero for the remaining $5n$ observations. The resulting coefficient estimates are consistent but not as efficient as the models that take into account the

---

[15]As noted above, the data set includes five rejected alternatives for each plant. To account for the implied clustering, each plant is represented by its own dummy variable. Thus, the conditional choice model has a distinct intercept for each plant.
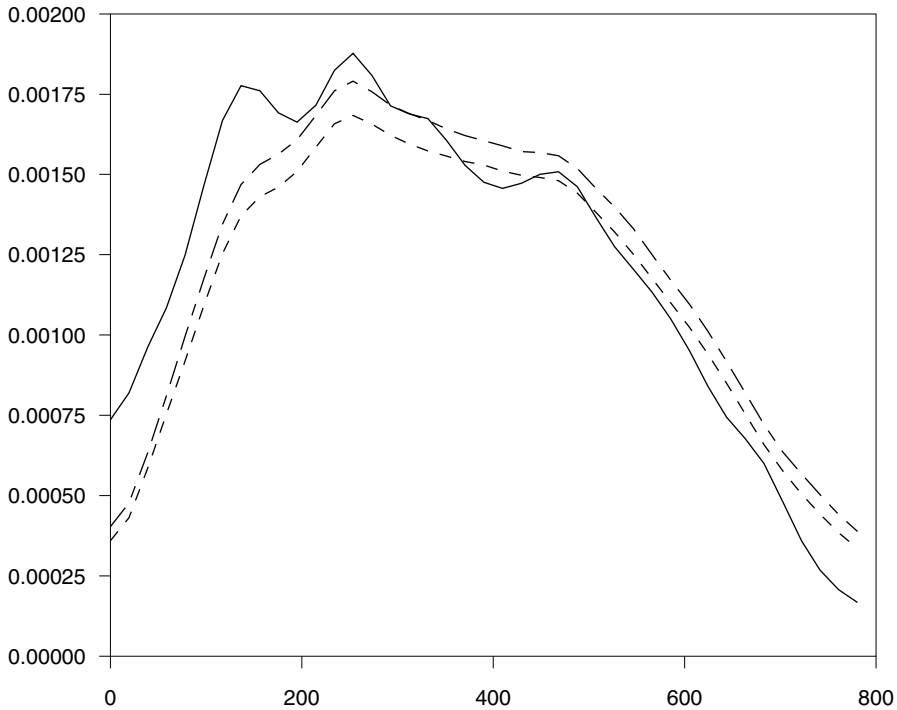
FIGURE 7: New Supplier Plant Density and Confidence Interval-
Estimated Probability of a New Plant.

clustering by plant group. However, the re-estimated logit model includes an intercept, and the coefficients can be used to calculate probabilities for *every* zip code.[16]

Figure 7 shows the actual K-density for new supplier plants and the bootstrap global confidence interval implied by the probabilities estimated using the model of new plant locations.[17] Comparing the confidence intervals across Figures 4–7, we see that a much lower degree of concentration is implied by comparing actual densities to the densities implied by the new plant logit model. The actual K-density function is still above the 95 percent global confidence interval in Figure 7, but it is much closer than was the case when the confidence intervals were based on zip code employment levels or uniform draws from all zip codes. Thus, the logit model has succeeded in explaining much of the tendency toward geographic concentration. Explaining the degree of concentration

---

[16]Apart from the intercept, the coefficients of the conditional logit model and the multinomial re-estimated model are nearly identical.

[17]The probabilities are based on the model without the variable indicating the number of assemblers within 100 miles.
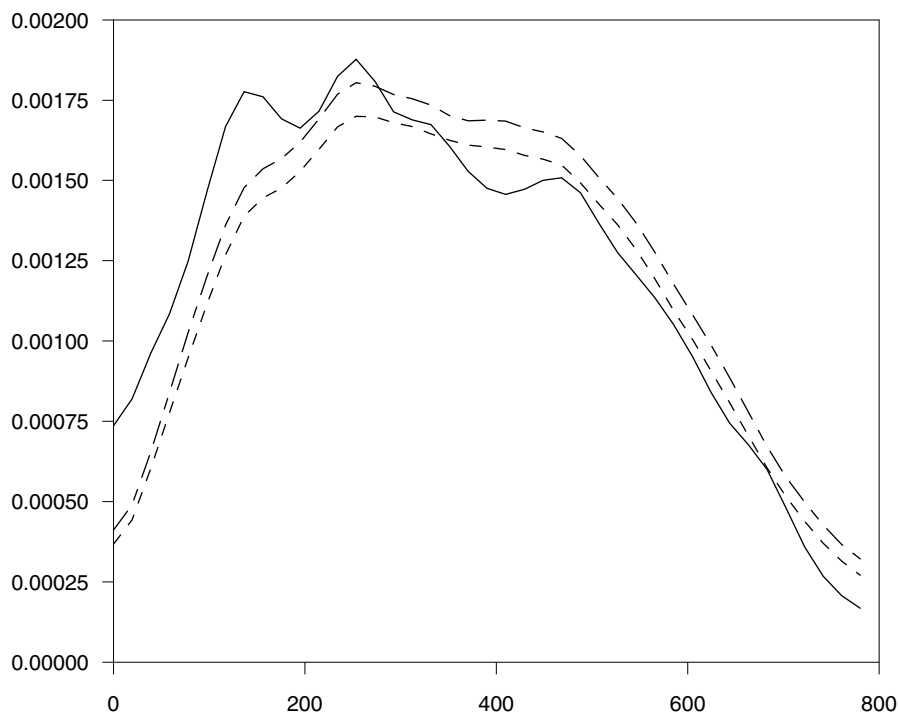
FIGURE 8: New Supplier Plant Density and Confidence Interval-
Estimated Probability of an Existing Plant.

further would require more explanatory variables or a model that explicitly
takes account of spatial autocorrelation.[18]

As shown in Figure 8, calculating confidence intervals for the *K*-density
function based on the existing plant logit probabilities produces a diagram
that is virtually identical to Figure 7. This result is not surprising since the
correlation between the two estimated sets of probabilities is 0.90. Figure 9
shows the global confidence intervals when we replace the estimated old-plant
logit probabilities with *actual* old-plant locations. To construct these confidence
intervals, we randomly draw samples of 431 locations from the 2,196 actual
old-plant locations. We then measure the distance of the 431 actual new plants
to the randomly drawn sample of locations. Aside from minor differences, the
resulting 95 percent global confidence interval contains the new-plant density
function. In other words, the distribution of new-plant distances is nearly the

---

[18]The literature on discrete choice models with spatial autocorrelation is still largely undevel-
oped. Relevant models include those proposed by Beron and Vijverberg (2004), Case (1992), LeSage
(2000), McMillen (1992), and Pinkse and Slade (1998). Currently, the models are only practicable
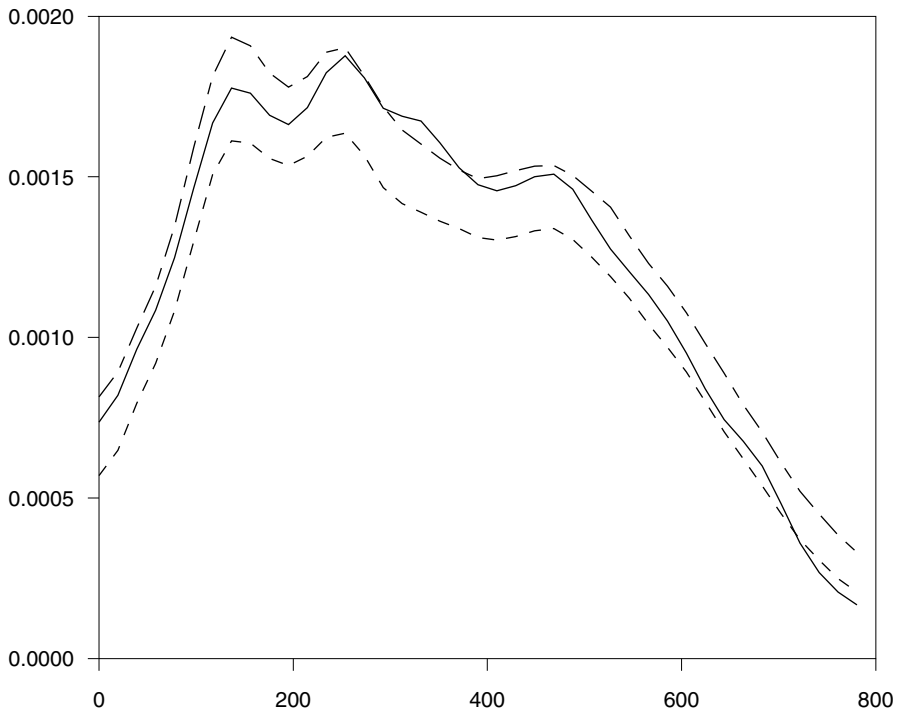for relatively small datasets because they involve inverting large weight matrices.

FIGURE 9: New Supplier Plant Density and Confidence Interval-
Actual Locations of Existing Plants.

same as what would be expected if new plants locations were simply drawn randomly from the sites of old plants. This result does not imply, of course, that new plants actually locate in the same sites as old plants. The importance of the result is that new plants show no additional tendency to cluster beyond the level of concentration of old plants. As the auto industry changed its orientation southward, the overall level of concentration remained essentially the same as before.

7. CONCLUSION

For the past century, the U.S. auto industry has been characterized by a small number of assembly plants and a large number of clustered supplier plants. Detroit remains the hub of the industry even as foreign plants have become more prominent. As American companies closed plants on the coasts and re-trenched toward the middle of the company, the industry has spread southward. The geographic distribution of auto supplier plants now displays a north-south orientation, with a concentration of plants along a corridor running from Detroit southward through Ohio, Kentucky, Tennessee, and into Alabama.

In this paper, we use a combination of nonparametric and parametric techniques to characterize the geographic distribution of auto supplier plants in the eastern United States. Using a nonparametric procedure developed by Duranton and Overman (2005), we find that auto supplier plants are much more concentrated than would be implied by random location choice. We then investigate the roots of this geographic concentration using parametric conditional logit models. We find that the location choices of the U.S. auto supplier plants are well explained by a small set of variables: the probability that a zip code has a plant is higher if the zip code has good highway access, is close to Detroit, and is near assembly plants. We also find that new supplier plants—those that have opened since 1991—are more likely to locate in zip codes that are near existing concentrations of supplier plants. Despite the recent change in the geographic orientation of the industry, both the nonparametric and parametric procedures suggest that the distribution of plants has not changed significantly over time. Although plant openings have been concentrated in the area south of Detroit, the new location pattern mimics the distribution of existing plants in the area.

Our results also suggest the usefulness of Duranton and Overman's (2005) procedure as a specification test for the conditional logit models. Although the logit models fit the data well, we find that plant locations are more concentrated geographically than is implied by the predicted logit probabilities. This result calls for the development of discrete choice models that explicitly take account of spatial clustering.

REFERENCES

Ben-Akiva, Moshe and Steven Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Beron, K. J. and Wim P. M. Vijverberg. 2004. "Probit in a Spatial Context: A Monte Carlo Analysis," in Luc Anselin, Raymond J. G. M. Florax, and Sergio J. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. New York: Springer, pp. 169–195.

Case, Anne C. 1992. "Neighborhood Influence and Technological Change," *Regional Science and Urban Economics*, 22, 491–508.

Duranton, Gilles and Henry G. Overman. 2005. "Testing for Localisation using Microgeographic Data," *Review of Economic Studies*, 72, 1077–1106.

———. 2006. "Exploring the Detailed Location Patterns of UK Manufacturing Industries using Microgeographic Data," manuscript.

Ellison, Glenn and Edward L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105, 889–927.

Harbour Consulting. 2003. *Harbour Report*.

Klier, Thomas and Daniel P. McMillen. 2006. "The Geographic Evolution of the U.S. Auto Industry," *Economic Perspectives*, 30, 2–13.

LeSage, James P. 2000. "Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models," *Geographical Analysis*, 32, 19–35.

McMillen, Daniel P. 1992. "Probit with Spatial Autocorrelation," *Journal of Regional Science*, 32, 335–348.

Pinkse, Joris and Margaret E. Slade. 1998. "Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models," *Journal of Econometrics*, 85, 125-154.

Rubenstein, James M. 1992. *The Changing U.S. Auto Industry — A Geographical Analysis*. London: Routledge.

Silverman, A. W. 1986. *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

Smith, Donald and Richard Florida. 1994. "Agglomeration and Industrial Location: An Econometric Analysis of Japanese-Affiliated Manufacturing Establishments in Automotive-Related Industries," *Journal of Urban Economics*, 36, 23–41.

Woodward, Douglas. 1992. "Locational Determinants of Japanese Manufacturing Start-ups in the United States," *Southern Economic Journal*, 58, 690–708.