LOCALIZATION OF INDUSTRY AND VERTICAL DISINTEGRATION

Thomas J. Holmes*

Abstract—Theory suggests that vertical disintegration should be greater in areas where industries localize. This paper provides some evidence that this implication is true for the U.S. manufacturing sector. Purchased inputs as a percent of the value of output is used as a measure of vertical disintegration. To measure the localization of industry, for each manufacturing plant the amount of employment in neighboring plants in the same industry is determined.

I. Introduction

IN HIS FAMOUS textbook, Marshall (1920) discussed a number of ideas about the phenomenon of localized industries. One of these ideas is that geographic concentration of an industry makes it possible for a host of specialized intermediate-input producers to emerge in areas where the industry is localized. A closely related idea, discussed by Stigler (1951) in his classic paper, is that concentration of industry may encourage *vertical disintegration;* i.e. the tendency to obtain inputs from specialized outside suppliers rather than making them within an integrated plant.[1]

Marshall's idea that localization facilitates the emergence of a wide variety of specialized suppliers has played a central role in a large body of theoretical work.[2] Stigler's theory has been the basis of several theoretical papers.[3] Less progress has been made in determining whether or not Marshall's and Stigler's ideas have widespread empirical relevance. A number of anecdotes and case studies illustrate the emergence of specialized suppliers and vertical disintegration for particular examples of localized industries.[4,5] But it is hard to say whether these examples illustrate a phenomenon that is rare or one that is common.

This paper presents some preliminary evidence of a link between localization of industry and vertical disintegration in the U.S. manufacturing sector. It considers census data on purchased inputs of manufacturing establishments. (This is

the value of intermediate goods purchased from outside suppliers, as opposed to intermediate goods produced internally.) Following Adelman (1955), the value of purchased inputs as a percent of total sales is used as a measure of vertical disintegration. Call this measure *purchased-inputs intensity*. The main finding is that establishments within an area where an industry is localized tend to be more vertically disintegrated in terms of this measure than establishments outside of the localization area.

This paper employs a novel method for handling the geographical data. A large body of literature measures the relationship between industry concentration in an area and various measures such as productivity, wages, or employment growth.[6] A challenge faced in this literature is how to deal with the arbitrary nature of the geographic boundaries upon which the data collection is based. Typically, studies have used employment in a city as a measure of industry concentration. With this approach, it is not clear whether San Francisco and Oakland should be treated as the same or different cities. This study avoids this difficulty by holding fixed the absolute size of the geographic unit (a circle with a radius of fifty miles).[7] Approximate longitude and latitude coordinates for all manufacturing establishments are obtained. For each manufacturing plant, I determine the neighbors of the plant; i.e. the other plants that are within fifty miles of the given plant.

The paper combines data on purchased inputs, aggregated to the level of locations, with data on employment, at the establishment level. With these data, the paper obtains estimates of the relationship between the purchased-inputs intensity of a plant and the level of employment of neighboring plants in the same industry. A central finding is that a plant with anywhere from 10,000 to 25,000 in own-industry neighboring employment has a purchased-inputs intensity that, on average, is three percentage points higher than a plant with fewer than 500 in own-industry neighboring employment. To put this difference in perspective, note that, in the manufacturing sector, purchased-inputs are on the order of 50% of sales for the average plant. So a three-percentage-point difference is a change from 50 to 53, an increase of 6%. This is the magnitude of the change in purchased-inputs intensity that is found when moving from a plant in an isolated area to a plant in a localized area. This estimate is not large, but neither is it negligible. It is also worth noting that this is the average change over the entire

[1] See also Hoover (1943) and Chinitz (1961).

[2] See, for example, Rivera-Batiz (1988), Krugman (1991), and Rodríguez-Clare (1996).

[3] See Perry (1989) for a survey.

[4] To begin, there is the famous example of the gun industry in nineteenth-century Birmingham (Allen, 1929). There are also studies of the aircraft industry in Southern California (Scott & Mattingly, 1989), the printed circuits industry in Southern California (Scott & Kwok, 1989), and a variety of industries concentrated in New York (Hall, 1959; Lichtenberg, 1960). All of these studies document the importance of specialized suppliers in industrial districts. (Additional references can be found in Scott (1983).)

[5] I should also mention that there has been empirical work that investigates some of Marshall's other ideas about why industries localize. In particular, Jaffe et al. (1993) present evidence on the enhanced information flows that can occur when industries localize.

[6] See, for example, Henderson (1986) and Glaeser et al. (1992).

[7] An alternative way to avoid the difficulty is considered by Ciccone and Hall (1996). They look at density; i.e., they divide employment levels of the geographic units by the land area of the geographic units. A problem with using density is that it does not take into account the fact that agglomeration effects can spill across the boundaries of the geographic units.

manufacturing sector; for many industries, the change is larger.

The results of this paper are relevant for the following question: How does a change in the scale of production at a location affect the organization of production at a location? Interest in this fundamental question dates back to Adam Smith. The positive correlation I find in the data between vertical disintegration and localization is consistent with the theory that an exogenous increase in the scale of industry leads to vertical disintegration.

The results of this paper are also relevant, albeit to a lesser extent, to a second question: Why do industries localize? In particular, is there any evidence that gains from being able to vertically disintegrate and purchase from a wide variety of local suppliers is a contributing factor for why some industries localize? My paper does not provide any direct evidence on this question. Nevertheless, the paper serves a useful purpose. If I had found that vertical disintegration were completely independent of localization, it would suggest that increasing the opportunity for vertical disintegration is probably not an important reason for industries to localize. Since I found otherwise, it holds open the possibility that it is a contributing factor, and it suggests that further work is warranted.

My results must be interpreted with caution. My finding of a positive correlation between localization and vertical disintegration is not proof of causality. I cannot tell what is causing what, or whether some third factor that causes localization and vertical disintegration induces a positive correlation between these variables in the data. Another reason for caution is that the measure of vertical disintegration I use has well-known inadequacies; the observed differences in purchased-inputs intensity that are found may signify something else besides differences in vertical disintegration. Given these difficulties, more research is needed to attack the question from different angles. This paper should be viewed as one small step in part of a larger research agenda of assessing the relevance of Marshall's and Stigler's insights for today's economy.

The rest of the paper is organized as follows. Section II discusses the basic idea of the paper and its limitations. Section III describes the data. Section IV describes the results. Section V considers an alternative approach that uses information about the location of specialized suppliers. Section VI concludes.

II.   The Basic Idea

The idea of this paper is very simple and can be illustrated with a stylized model of the pantyhose industry. Suppose there are three stages of production: spinning fiber into yarn, knitting yarn into hose, and finishing the hose (i.e., bleaching and dyeing it). Suppose the cost of each stage of production is $0.50 per pair of hose and, for simplicity, that the fiber used in the first stage is a free good. Assume that the competitive price of a pair of pantyhose is $1.50 (the sum of the cost of all three stages of production).

The pantyhose industry happens to be heavily concentrated in North Carolina. Suppose there exists a pantyhose factory outside of North Carolina that is relatively isolated in the sense that there are no other pantyhose factories within hundreds of miles. The theories developed by Marshall, Stigler, and others suggest that this plant is likely to be more vertically integrated than a plant in North Carolina. Suppose the plant is completely integrated and does all three stages. Then the cost of its purchased inputs is 0 (remember fiber is free), and the value of its sales is $1.50, so purchased-inputs as a percent of sales is 0. If the plant purchases its yarn but does knitting and finishing itself, then purchased inputs is $0.50 and purchased-inputs intensity (PII) is $\frac{.50}{1.50} = \frac{1}{3}$.

Now consider plants in North Carolina where the industry is concentrated. Theory suggests that vertical disintegration should be prevalent here. Suppose production occurs in three specialized plants. The first plant spins yarn from the free fiber; its PII is 0. The second plant buys yarn for $0.50 and sells unfinished hose for $1.00; its PII is $\frac{1}{2}$. The third plant buys unfinished hose for $1.00 and sells finished hose for $1.50; its PII is $\frac{2}{3}$. This example illustrates a well-known inadequacy of PII as a measure of vertical disintegration. (Adelman, 1955). All three plants are vertically integrated to the same degree as they all undertake one stage of the production process. But they vary in PII because they vary in how far downstream they are in the production process. Despite this inadequacy of this measure, it is nevertheless the case that the sales-weighted average PII of the plants in North Carolina (equal to $\frac{1}{2}$) is greater than the PII of the plants outside of North Carolina (equal to 0 in the fully-integrated case, and $\frac{1}{3}$ in the partially integrated case).[8] So by looking at this measure, we draw the correct conclusion for this example that the plants in North Carolina are less vertically integrated than the plants outside of North Carolina.

A fundamental issue that must be addressed in the approach taken here is this: if there are efficiency advantages to localizing the production of a particular industry, why would anyone ever locate a plant in an isolated location without any own-industry neighbors?

One possible answer to this question is that a plant in an isolated location may have some advantage that offsets the absence of agglomeration economies. There might exist a willingness to pay extra for a locally-produced good. The higher price a local plant might receive might enable the plant to break even despite its higher costs. Analogously, there may exist some industry-specific supply factors at an isolated location that might attract a plant (e.g., there might be some people in Montana who have always wanted to work in a pantyhose plant). In either case, isolated plants will tend be more vertically integrated and have a lower PII

---

[8] The $\frac{1}{2}$ figure is obtained by weighting the North Carolina plants by the value of output, as will be the case in the empirical analysis. Even in the unweighted case, the mean of the North Carolina plants at $\frac{7}{18}$ exceeds the PII of the plants outside North Carolina.

than plants with many neighbors, so using my procedure gets the right answer.

Another possibility is that production in isolated areas may serve a different function than production in localized areas. There are at least two ways that function can differ across locations.

One difference in function between locations is a difference in the kinds of products produced. For example, as argued by Lichtenberg (1960), plants in concentrated areas (like the garment industry in New York) might make fashion-oriented goods (e.g., evening gowns), while plants in isolated areas might specialize in the production of standardized goods (e.g., nurse uniforms). There are reasons to believe that isolated plants making standardized goods would tend to be more vertically integrated than plants in the concentrated areas. By looking at *PII,* we should pick this up, so my procedure should get the right answer here. But other examples can be constructed for which things don't work out so well. Suppose plants at two locations differ in the quality of the final product, with quality defined as how fast a given unit of product is pushed through the production process. Plants at two different locations may actually have the same level of vertical integration; e.g., they may each do a single step only. But, if one plant does the step slower (to make a higher quality final product), purchased inputs will be a smaller share of the value of output. This second example shows that, if different locations specialize in different kinds of products, it is possible that my procedure will give the wrong answer. The procedure might suggest that plants at different locations differ in the degree of vertical integration even if those differences do not exist.

A second difference in function is that different locations may specialize in different stages of the production process. For example, for many years the U.S. auto industry was organized with parts plants concentrated in Michigan and assembly plants spread throughout the county.[9] There can be some savings in transportation cost to having the final stages of production near final consumers. Suppose that parts plants and assembly plants are vertically integrated to the same degree: they both undertake one task. The assembly plants will nevertheless have a higher *PII* than parts plants because they are further downstream. In this case, a comparison of isolated plants (the assembly plants) with plants with neighbors (the parts plants in Michigan), would reveal that the isolated plants have a higher *PII* than plants with neighbors. To the extent this factor is operative, examining the difference in *PII* between plants with neighbors and plants without neighbors will understate the extent to which production in the localized area is more disintegrated.

III.   The Data

This section discusses the data. It begins by describing the variables collected by the census. It then explains the procedure I used to handle the geographic nature of the data set.

A.   The Variables Collected by the Census

The data is from the 1987 Census of Manufactures. This data is collected at the establishment level (a factory or plant at a particular location) as opposed to the firm level. Each establishment was asked to report information about the activity of the plant in 1987.

The key census variable for my purposes is *purchased inputs.*[10] This variable is the value of intermediate goods obtained from other establishments. According to the census, "It includes the cost of materials or fuel consumed, whether purchased by the individual establishment from other companies, transferred to it from other establishments of the same company, or withdrawn from inventory during the year." This purchased-inputs variable includes any freight charges incurred in shipping the inputs to the establishment. The variable does not include the value of any business services obtained by the establishment. This omission is unfortunate for my purposes, because many business services are likely to be exactly the kind of locally produced intermediate input that producers in localized areas will have greater access to than producers in isolated areas. Because the purchased-inputs variable does not include business services, my estimates are likely to understate the extent to which production in localized areas is more disintegrated than production in isolated areas.

Two other census variables are important in the analysis. The variable *employment* is the number of full- and part-time employees of the establishment in mid-March, 1987. The variable *output* is the receipts from sales of products that left the plant in 1987. (The census calls this variable *value of shipments.*)

B.   Construction of the Data Set

It is useful to begin by explaining what the procedure would be if ideal data were available. Ideal data would consist of the records for each establishment in the 1987 Census of Manufacturers. With these records, it would be possible to pinpoint the exact location (the longitude and latitude coordinates) of each establishment. Suppose a circle with a fifty-mile radius were drawn around each plant. Define a neighboring establishment to be any establishment that is located within this circle, i.e., within fifty miles as the crow flies. Next, calculate purchased inputs as a percent of output at each plant. The data are now ready to be used to determine the relationship between purchased-inputs intensity and the amount of neighboring employment.

The ideal data set just described is not publicly available. However, a substantial amount of data is distributed on a CD-ROM. This publicly available data permits an analysis

---

[9] See Rubenstein (1992).

[10] The census actually refers to this variable as *cost of materials.* For my purposes, it is convenient to rename the variable *purchased inputs.*

that is surprisingly close to the one described above. Setting up this data involves two steps. The first step determines the neighboring employment of establishments. The second step squeezes out the maximum information possible about purchased inputs from the publicly available data.

C. Neighboring Employment

The Location of Manufacturing Plants file reports the number of establishments in each of seven employment size categories by four-digit industry and county. This is an establishment-level data set. For each one of the 368,896 manufacturing establishments existing in 1987, the data set provides information about the industry, the employment size, and the location of the establishment. Let i_e denote the four-digit industry of establishment e, where e indexes the 368,896 establishments in the data set. The employment variable and the location variable require more discussion.

The seven size categories for the number of employees are 1–19, 20–49, 50–99, 100–249, 250–499, 500–999, and 1,000 or more. I converted this categorical variable to an estimate of the employment for each establishment. For each size class, I calculated the mean number of employees across all the establishments within the size class. (This information is available from aggregate data.) For example, in the 1–19 size class, the mean of employment is 6 while, in the 20–49 size class, the mean of employment is 31. For each establishment e, define $\widehat{employment}_e$ to be the mean number of employees within the establishment's size class. For example, if it is in the 1–19 class, $\widehat{employment}_e = 6$, and, if it is in the 20–49 class, $\widehat{employment}_e = 31$.[11]

The location information is the county that the establishment is in. I obtained the longitude and latitude of the geographic center of each county and used these coordinates to approximate the location of each establishment. Let $\widehat{longitude}_e$ and $\widehat{longitude}_e$ be the approximate longitude and latitude of establishment e.

I define the *neighborhood* of a plant as the county that the plant is in, as well as the counties with centers that are within fifty miles (as the crow flies) of the center of the county that the plant is in. This neighborhood will approximate to some extent a circle with a radius of fifty miles drawn around the location of the plant. To get some idea of how good this approximation is, consider figure 1. To construct this figure, I selected some arbitrary counties to serve as *center* counties (counties illustrated in gray) and then determined the neighboring counties located within fifty miles of the center (counties illustrated in black). Outside the western states, the neighborhoods look something like circles with a radius of fifty miles. In states like Nevada and Arizona, the neighborhoods look nothing like such circles because the counties are so large. (Several of the counties in these states are bigger

than New Jersey.) Fortunately (for my purposes), there are relatively few establishments in states like Nevada.

For each establishment e, define $neighbor_e^{own}$ to be the total employment of neighboring establishments (those located within fifty miles) that are in the same four-digit industry as establishment e. This is own-industry neighboring employment. Note this variable does not include the employment of establishment e: it is the employment outside of this factory located nearby. The variable is calculated by using the estimates for the employment of each establishment described above. A second variable, $neighbor_e^{related}$, is related-industry neighboring employment. This is employment of neighboring establishments within the same two-digit industry as establishment e, but outside of the same four-digit industry. A third variable, $neighbor_e^{other}$, is other-manufacturing neighboring employment. This includes all neighboring manufacturing employment except for employment within the same two-digit industry as establishment e. All neighboring manufacturing employment of establishment e is in one of the three mutually exclusive categories: own-industry, related-industry, or other-manufacturing.
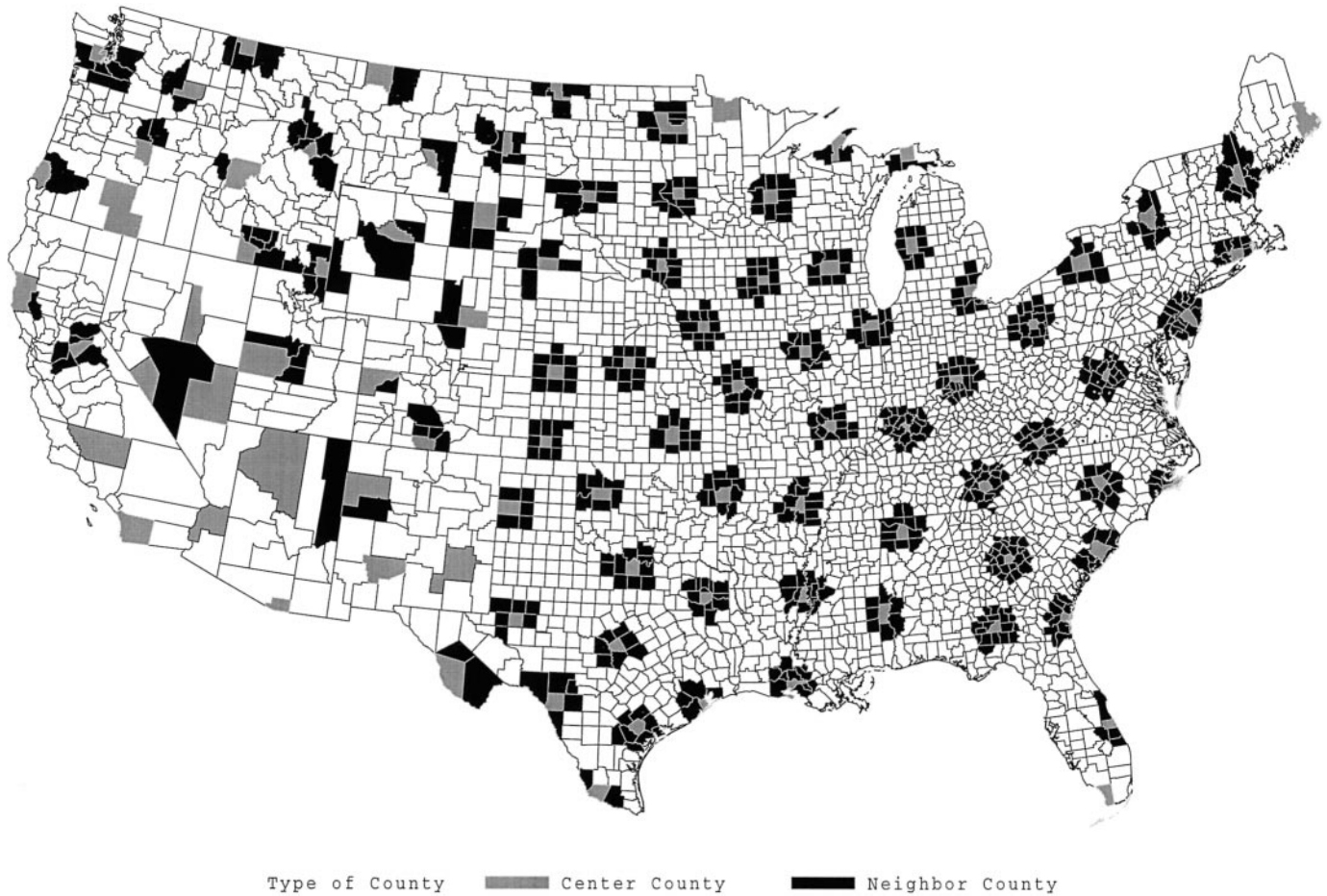
D. Purchased Inputs and Output

The variables *purchased inputs* and *output* are not publicly available at the establishment level. The census publishes aggregates for the entire United States and for selected areas. These areas can be states, metropolitan statistical areas (MSAs), and counties. Selected areas vary for different industries. For example, for the creamery butter industry (SIC = 2021), this data is available for the state of Wisconsin as well as for the United States, but for no other geographic area. In contrast, for the commercial printing industry (SIC = 2752), this data is available for over 200 geographic units.

For each industry, I used the data available from the census to partition the set of counties in the United States into nonoverlapping areas for which it is possible to determine aggregate purchased inputs and output for the establishments in the areas. For example, there are two areas for the creamery butter industry: the first is Wisconsin and the second is all of the United States except Wisconsin. (The data for the latter is obtained by subtracting the Wisconsin totals from the U.S. totals.) For some industries, I created areas that consisted of the balance of the counties in a state (when I had data on the state and some counties in the state) and other areas that consisted of the balance of the counties in a MSA (when I had data on the MSA and some of the counties in the MSA).

Let ia index a particular industry i in a particular area a. Let *purchased inputs$_{ia}$* and *output$_{ia}$* denote the totals for these variables obtained by summing over the levels of the variables for all establishments in industry i and area a.

---

[11] Mean employment in the "1,000 or more size" class varies substantially across industries. For this size class, the mean within the two-digit industry was used.

Figure 1.—A Sample of Counties and Their Neighborhoods

Type of County   Center County   Neighbor County

E. The Industry/Area Data Set

The industry/area data set is obtained by combining the purchased-inputs and output data for each industry/area described in subsection D with the establishment-level data on neighboring employment described in subsection C.

The establishment-level data has to be aggregated up to the level of an industry/area in some way. Here is the procedure that I used. Note first that purchased-inputs intensity (*PII*) in an industry/area can be written as the weighted sum of the *PII* at the establishment level,

$$PII_{ia} \equiv \frac{purchased\ inputs_{ia}}{output_{ia}}$$
$$= \sum_{e \in ia} w_{eia} \cdot \frac{purchased\ inputs_{eia}}{output_{eia}}, \tag{1}$$

with the weights given by the establishment share of industry/area output,

$$w_{eia} \equiv \frac{output_{eia}}{output_{ia}}. \tag{2}$$

This suggests an aggregation of the neighboring employment data that uses establishment output share as the weight. Unfortunately, I do not observe establishment output share. However, an establishment's output share will be related to its employment share, and I can construct an estimate of establishment employment share by

$$\hat{w}_{eia} \equiv \frac{\widehat{employment_{eia}}}{\widehat{employment_{ia}}}. \tag{3}$$

Now, define average own-industry neighboring employment in industry/area *ia* by

$$neighbor_{ia}^{own} \equiv \sum_{e \in ia} \hat{w}_{eia} \cdot neighbor_{eia}^{own}. \tag{4}$$

Average related-industry employment $neighbor_{ia}^{related}$ and average other-manufacturing employment $neighbor_{ia}^{other}$ are defined in the analogous way.

Table 1 presents some summary statistics of the industry/area data set. The first part of the table reports the distribution of number of areas across industries. There are 26

TABLE 1.—INDUSTRY/AREA DATA SET SUMMARY STATISTICS

A. Distribution of Number of Areas Across Industries

Number of Areas	Number of Industries With This Number of Areas
1	26
2–5	110
6–10	99
11 or More	224
Subset of Industries With Multiple Areas	
Number of Industries	433
Number of Industry/Areas	8,096

B. Summary Statistics of Selected Area Variables

Description	Notation	Mean	Std. Dev.	Min.	Max.
Number of Establishments	$number_{ia}$	44.2	102.7	1	4226
Employment (thousands)	$employment_{ia}$	2.1	5.7	0.01	216.5
Purchased-Inputs Intensity	$100 \times \dfrac{purchased\text{-}inputs_{ia}}{output_{ia}}$	47.3	14.0	1.9	102.9
Neighboring Employment Measures (thousands of employees)					
Own-Industry	$neighbor_{ia}^{own}$	2.3	4.7	0	74.1
Related-Industry	$neighbor_{ia}^{related}$	25.4	31.6	0	195.2
Other-Manufacturing	$neighbor_{ia}^{other}$	315.3	320.3	0	1178.9

industries with only one area. (The census reports data only for the entire United States for these industries.) These industries will be deleted in the analysis below, because there is no cross-sectional variation across areas within these industries. After this deletion, 433 industries remain. Approximately half of these industries have two to ten areas and the other half have eleven or more.

One possible concern about the aggregation here is that industries that are unconcentrated may be overrepresented in the data set. It is the case in these data that industries for which there are a large number of areas on average tend to be less geographically concentrated than industries for which there are few areas.[12] However, the empirical analysis will use establishment employment to weight the industry/area observations so that unconcentrated industries will not have a disproportionate effect on the results, even though they account for a disproportionate amount of areas.

IV.   Empirical Results

This section examines the relationship between purchased-inputs intensity and geographical concentration of industry. The first part of this section presents some preliminary numbers calculated in a simple way that avoids the complicated treatment of the geographical data explained in the previous section. The second part of this section uses the industry/area data set to examine the issue.

A.   Some Preliminary Numbers

Section II stated a simple prediction. For the pantyhose industry, average purchased-inputs intensity (*PII*) of plants

in North Carolina (where this industry is heavily concentrated) should be higher than the average *PII* of plants outside North Carolina. This subsection investigates whether this simple prediction is true for the pantyhose industry and whether a similar relationship tends to hold for other geographically concentrated industries like the pantyhose industry.

As discussed below, the procedure I use in this subsection makes sense only when it is applied to industries that are heavily concentrated in a particular state. To determine which industries are geographically concentrated, I employ the measure recently developed by Ellison and Glaeser (1997).[13] Table 2 lists the forty most-geographically concentrated industries according to this index, ranked in descending order of the index.[14] In this table, the center of the industry is defined to be the state with the highest share of employment in the industry.

The pantyhose industry (Women's Hosiery Except Socks) is fifth on the list. North Carolina has a 62% share of the national employment in this industry. The *PII* is 53% in North Carolina and only 40% outside North Carolina. To the extent that purchased-inputs intensity is a proxy for vertical disintegration, there is greater disintegration inside of North Carolina than outside the state.[15]

The pattern that the *PII* is higher in the center state than outside the center state is true for 31 of the industries

---

[12] For example, the mean Ellison-Gleaser index (an index of concentration mentioned below) for industries with eleven or more areas is half the mean index for industries with ten or fewer areas.

[13] I use their estimates of the index that they report in the appendix of the working paper version of their paper.

[14] More precisely, the top forty industries for which the data required by the table are available. This eliminates fourteen industries.

[15] A tendency for pantyhose plants in North Carolina to be vertically disintegrated is consistent with what I have learned from talking to an industry source and examining census data on product shipments. It appears that the movement of unfinished pantyhose across establishments is more important in North Carolina area than outside this area.

TABLE 2.—PURCHASED-INPUTS INTENSITY FOR LOCALIZED INDUSTRIES

	Four-Digit Industry	Center	Center Share of Employees (%)	Purchased-Inputs Intensity (%)	
				Center	Rest of U.S.
2371	Fur Goods	New York	77	71	60
2084	Wines, Brandy, Brandy Spirits	California	78	58	54
2252	Hosiery, n.e.c.	North Carolina	66	**52**	**52**
3533	Oil Field Machinery	Texas	66	47	42
2251	Women's Hosiery Except Socks	North Carolina	62	53	40
2273	Carpets and Rugs	Georgia	62	73	62
2429	Special Product Sawmills, n.e.c.	Washington	59	59	55
3961	Costume Jewelry	Rhode Island	54	40	36
2895	Carbon Black	Texas	39	56	53
2874	Phosphatic Fertilizers	Florida	54	**68**	**70**
2061	Raw Cane Sugar	Hawaii	32	**46**	**71**
2281	Yarn Spinning Mills	North Carolina	49	**58**	**61**
2034	Dehydrated Fruits	California	59	57	41
2083	Malt	Wisconsin	43	69	69
2221	Weaving Mills, Synthetics	South Carolina	41	57	55
2284	Thread Mills	North Carolina	46	65	60
2282	Throwing and Winding Mills	North Carolina	40	79	68
2257	Circular Knit Fabric Mills	North Carolina	44	65	57
2262	Finishing Plants Synthetics	South Carolina	39	76	70
2436	Softwood Veneer and Plywood	Oregon	38	62	59
2296	Tire Cord and Fabric	Georgia	45	64	59
3465	Automotive Stampings	Michigan	40	**54**	**58**
3861	Photographic Equipment	Massachusetts	9	38	32
2211	Weaving Mills, Cotton	North Carolina	33	53	53
2387	Apparel Belts	New York	46	50	45
3552	Textile Machinery	South Carolina	31	45	44
2397	Schiffli Machine Embroideries	New Jersey	36	52	40
3716	Motor Homes	Indiana	40	**72**	**73**
2021	Creamery Butter	Wisconsin	29	93	90
3171	Women's Handbags and Purses	New York	42	44	42
3572	Computer Storage Devices	California	48	51	50
2824	Organic Fibers, Noncellulosic	South Carolina	26	54	50
2046	Wet Corn Milling	Iowa	28	**52**	**58**
2044	Rice Milling	Arkansas	29	85	61
2022	Cheese, Natural and Processed	Wisconsin	35	83	78
2512	Upholstered Household Furn.	North Carolina	34	51	50
3711	Motor Vehicles and Car Bodies	Michigan	37	74	73
2261	Finishing Plants, Cotton	North Carolina	19	**57**	**57**
3743	Railroad Equipment	Pennsylvania	39	**42**	**56**
2258	Lace and Warp Knit Fabric Mills	North Carolina	33	60	59
	Mean Values			59.6	56.6

considered. (The nine exceptions are in bold print.) Hence, the predicted pattern holds for an overwhelming majority of the industries. To give some sense of the statistical significance of this finding, consider a null hypothesis that, for a randomly selected industry, the probability is one-half that the *PII* is greater inside the center than outside it. Then, when sampling forty industries, the probability of drawing 31 or more industries in which the measure is greater in the center is only 0.0003. (This is the probability of drawing 31 or more heads in 40 flips of a fair coin.) The average difference across the forty industries between the *PII* in the center and the *PII* outside the center is 3.0 percentage points. Under the null hypothesis that these forty industries are drawn from a universe where the average difference is zero, the probability of drawing an average difference bigger than 3.0 is 0.01.

This procedure does not make sense when applied to industries that are diffuse. For such industries, big states such as California will tend to have the largest share of national employment of the industry. When a state has a large share, it by no means implies that the industry is particularly concentrated in the state, in the sense of plants in the state having relatively many neighboring plants.

The procedure has limitations even when applied to industries that are concentrated. The procedure lumps together all the production outside of the center state and some of this production might be at concentrated locations. Analogously, the procedure lumps together all the production inside the center state, and, if the state is big, there may be substantial variations in concentration across the state. All of these limitations lead me to consider analysis of the industry/area data set where these difficulties are avoided.

B.   The Findings with the Industry/Area Data Set

I now consider various regression models to determine the relationship between purchased-inputs intensity (*PII*) and

TABLE 3.—PURCHASED-INPUTS INTENSITY AND NEIGHBORING EMPLOYMENT COEFFICIENT ESTIMATES FOR REGRESSION MODELS
(NEIGHBORING EMPLOYMENT IN 1,000s)

| | Estimated Coefficients for Linear and Cubic Models | | | | | | | | | |
| | Own-Industry | | | Related-Industry | | | Other-Manufacturing | | | |
Model	Level (1,000)	Level Squared	Level Cubed	Level (1,000)	Level Squared	Level Cubed	Level (1,000)	Level Squared	Level Cubed	R^2
1	0.04 (0.01)	—	—	—	—	—	—	—	—	0.001
2	0.04 (0.02)	—	—	0.000 (0.003)	—	—	—	—	—	0.001
3	0.05 (0.02)	—	—	0.009 (0.004)	—	—	−0.0013 (0.0004)	—	—	0.002
4	0.35 (0.07)	−0.013 (0.003)	0.00012 (0.00004)	−0.054 (0.023)	0.0009 (0.0003)	−3.0E-6 (1.2E-6)	0.0025 (0.0032)	−1.4E-5 (0.7E-5)	1.0E-8 (0.4E-8)	0.006

| | Estimated Coefficients for Size-Class Model | | | | | | | | | | | | | | |
| | Own-Industry | | | | | Related-Industry | | | | | Other-Manufacturing | | | | | |
Model	0–0.5	0.5–2.5	2.5–10	10–25	25+	0–2.5	2.5–10	10–25	25–100	100+	0–10	10–25	25–100	100–500	500+	R^2
5	x	0.6 (0.2)	0.2 (0.3)	1.4 (0.4)	2.0 (0.6)	—	—	—	—	—	—	—	—	—	—	0.003
6	x	1.2 (0.3)	1.4 (0.4)	3.1 (0.6)	3.7 (0.8)	x	0.8 (0.4)	0.8 (0.5)	0.1 (0.6)	1.0 (0.7)	x	1.8 (0.9)	0.9 (0.6)	−0.2 (0.7)	−1.0 (0.7)	0.009

own-industry neighboring employment ($neighbor^{own}$) in the industry/area data set.

My procedure allows for industry fixed effects by differencing the left-side variable and the right-side variables by industry means. The procedure regresses the average level of the *PII* in an industry/area (differenced from the industry mean) against the average levels of neighboring employment variables (also differenced from industry means). I use weighted least squares with the weights given by the number of establishments in an industry/area with a correction to take into account the asymmetry of establishment size among establishments within an industry/area. (This procedure is explained in the appendix.) The motivation for using this procedure is that, under certain assumptions, it identifies what the parameter estimates would be if I were to run the same regressions with establishment-level data (as opposed to the industry/area-level data that is actually used).

Table 3 presents the estimated coefficients for six alternative models. Model 1 is a simple linear model in which *PII* is regressed against $neighbor^{own}$. All the neighboring employment levels in the table are denoted in units of 1,000 employees. The estimated coefficient on $neighbor^{own}$ in Model 1 is 0.04. Under the assumptions discussed in the appendix, the estimated coefficient can be given the following interpretation. In a cross section of establishments within an industry, the conditional expectation of purchased inputs as a percent of output increases by 0.04 percentage points with a 1,000 increase in own-industry neighboring employment. Note that, while the conditional mean depends upon $neighbor^{own}$ in a statistically significant way, only a small fraction of the within-industry variation in *PII* can be accounted for by variations in $neighbor^{own}$. (The R^2 is only 0.001.)

Models 2 and 3 consider what happens in the linear model when related-industry neighboring employment and other-manufacturing employment are added as additional variables on the right-hand side. Adding these additional regressors does not have a big effect on the coefficient estimate for own-industry neighboring employment.

There are a priori reasons to believe that the relationship between *PII* and $neighbor^{own}$ employment should be concave. An earlier version of this paper presents a model in which this is the case. If nothing else, the fact that *PII* is bounded between 0 and 100 suggests it is useful to consider a nonlinear relationship. Models 4, 5, and 6 are three alternative nonlinear models. Before discussing the results, it is useful to discuss the relevant range of the data. The median establishment has an own-industry neighboring employment of 0.9 (thousand). Five percent of establishments have $neighbor^{own}$ equal to zero. At the other end of the distribution, 5% have $neighbor^{own}$ equal to 19,000 or more.

Model 4 estimates a cubic equation.[16] The estimate of the relationship between *PII* and $neighbor^{own}$ is very different from the linear model. The slope at 0 in this case is 0.35, seven times the slope of 0.05 in the linear model. Given the negative estimate of −0.013 on the squared term, the relationship is concave for small $neighbor^{own}$. For $neighbor^{own}$ in the range of 0 to 19,000 (a range that contains 95% of all establishments), this cubic function is strictly increas-

[16] To estimate this model, I begin with the establishment-level data set. I then square and cube the establishment-level neighboring employment variables. Next, I aggregate to the area level by taking the weighted average as in equation (4). Note that, because of the nonlinearities here, this procedure is different from what I would get if I first aggregated to the area level, and then squared and cubed the variables. Given the assumptions made in the appendix, the approach I am taking is the appropriate way to go. I do an analogous thing with models 5 and 6.

ing and strictly concave. The estimate of this cubic equation implies that, if we compare two establishments, one with *neighbor^{own}* equal to 0 (the fifth-percentile establishment), the other with *neighbor^{own}* equal to 19,000 (the 95th-percentile establishment), with the other variables held fixed, the *PII* of the second plant is higher by 2.9 percentage points. This is a substantially bigger effect than what I get in the linear model, where the expected difference is approximately one percentage point ($0.95 = 19 \times 0.05$).[17]

Models 5 and 6 consider an alternative nonlinear model in which the relationship is assumed to depend upon size classes of neighboring employment. One reason for considering such a model is that the estimated coefficients are easier to interpret than the coefficients of the cubic equation. To estimate models 5 and 6, each establishment was classified into one of five groups based on the establishment's own-industry neighboring employment (in thousands): *0–0.5, 0.5–2.5, 2.5–10, 10–25,* and *25+.* On the basis of this classification, four dummy variables were constructed with the *0–0.5* group being the excluded group. Thus, if an establishment is in the *0–0.5* group, the four dummy variables are zero; if an establishment is in the *0.5–2.5* group, the first dummy variable is 1 and the remaining three are zero; and so on. Analogous groupings were constructed for related-industry and other-manufacturing neighboring employment. These various groupings are displayed in the bottom of table 3.

Model 5 includes only the own-industry variables. Model 6 adds to this the related-industry and other-manufacturing variables. Since model 6 is comparable to model 4, I discuss it first.

In model 6, the coefficient estimate on the *0.5–2.5* size class dummy is 1.2. The interpretation of this coefficient is that the average difference in *PII* between establishments, with 500 to 2,500 in own-industry neighboring employment, and those, with less than 500, is 1.2. In calculating these differences, the industry of the establishments as well as the related-industry and other-manufacturing neighboring employment groups are held fixed. Now, consider the remaining coefficients on the own-industry dummy variables for model 6. These coefficients are all statistically significant from zero. Furthermore, they monotonically increase as we move across the row, from 1.2 to 1.4 to 3.1 to 3.7.

The estimated coefficients in model 5 are approximately half of the corresponding estimates in model 6. Nevertheless, the differences between the *0–0.5* category and the *10–25* and *25+* categories remain highly statistically significant.

I now briefly discuss the coefficients on the related-industry and other-manufacturing variables. In model 6, the average *PII*s for establishments in the related-industry size classes above 2,500 are slightly bigger than the *PII* of the *0–2.5* group. However, these differences are not statistically

significant. In model 4, the cubic model, the estimated difference between a plant with a positive *neighbor^{related}* and one with *neighbor^{related}* = 0 is actually slightly negative, for most of the relevant range of *neighbor^{related}*.[18] These results suggest that neighbors in related industries (i.e., ones with the same two-digit but a different four-digit SIC code) may not be close substitutes for neighbors in the same industry (ones with the same four-digit SIC code).

The estimates of the effect of other-manufacturing neighbors are similar in models 4 and 6. Plants with *neighbor^{other}* in the range from 10,000 to 100,000 on average have *PII* values that exceed those of isolated plants with *neighbor^{other}* less than 10,000. However, plants in the most populous areas actually tend to have a slightly lower *PII* than the most-isolated plants. This is surprising, because one might expect that urbanization effects alone will would lead to vertical disintegration.

V.   Evidence from the Location of Specialized Suppliers

This paper looks for evidence that plants in agglomerations are less vertically integrated than plants outside of agglomerations. The strategy used up to this point in the paper is to see whether or not plants in agglomerations tend to use purchased inputs more intensively than do isolated plants.

An alternative strategy is to examine the location of specialized suppliers. Theory suggests that specialized suppliers emerge in agglomerations. According to the theory, plants in agglomerations tend to outsource various stages of production to local, specialized suppliers, while plants not in agglomerations tend to undertake these stages of production as vertically-integrated establishments. Theory implies that a disproportionate amount of specialized suppliers will emerge at locations where an industry agglomerates.

While this second strategy has promise, two data limitations make it a difficult one to pursue. First, the census does not, in general, distinguish between specialized suppliers and vertically-integrated establishments when classifying establishments into industries. Establishments that undertake two stages of production routinely get lumped into the same industry as establishments that undertake one stage of production. Second, linking industries by vertical stage of production is a complicated project.

I have found one industry, the textile industry, for which these data limitations do not arise. In this section, I begin an exploration of how one might examine the location patterns of specialized suppliers, using data from this industry. I find some evidence that the location patterns of specialized suppliers in the textile industry is consistent with the earlier results about purchased-inputs intensity.

The census three-digit industry SIC 226 consists of plants that dye and finish textiles. The census is explicit that this industry consists of specialized plants. These are ''establish-

---

[17] Another way to allow for a concave relationship is to use $\log(1 + neighbor^{own})$ as the functional form. In this case, the estimate of the change from 0 to 19,000 is 2.2 percentage points.

[18] For *neighbor^{related}* in the range from 0 to 100,000 (which contains 95% of all establishments), the difference ranges from 0 to -1.

TABLE 4.—FRACTION OF COUNTY NEIGHBORING TEXTILE EMPLOYMENT IN FINISHING BY NEIGHBORING TEXTILE-EMPLOYMENT SIZE CLASSES

| | All Counties | By Neighboring Textile-Employment Size Classes | | | | | |
		1–99	100–499	500–999	1,000–4,999	5,000–9,999	10,000 and Above
Mean	0.096	0.163	0.057	0.056	0.052	0.096	0.091
Median	0.018	0.000	0.015	0.011	0.008	0.060	0.080
Number of Counties	2,390	776	479	261	471	128	275

ments primarily engaged in finishing purchased . . . fabrics, or finishing such fabrics on a commission basis,'' (U.S. Bureau of Census (1987, p. 91)). Integrated plants that finish fabric that they make themselves are excluded. In this section, the *textile industry* is defined as all establishments in the two-digit SIC 22 (Textile Mill Products), and *specialized finishing plants* are the subset of the textile plants that are in SIC 226.

The first question I consider is whether the results on purchased-inputs intensity that I obtained for the manufacturing sector as a whole apply for the textile industry in particular. I reran the regressions reported in table 3 with just data from the textile industry, and the results are similar to results for all of manufacturing. Consider, for example, the simplest regression of *PII* on own-industry neighboring employment (where the own industry is all textile plants, SIC 22). The estimated coefficient is 0.08 with a standard error of 0.03. This is larger, but of the same order of magnitude, as the estimate of 0.04 in Table 3 for the manufacturing sector as a whole. The estimated coefficient changes little whether or not I also include other manufacturing in the regression and whether or not I control for four-digit fixed effects.

The second question I consider is whether a disproportionate amount of specialized finishing plants are at locations where the textile industry agglomerates. The best way to do this is not obvious. In this section, I consider a method that is in the same spirit as my method for looking at purchased inputs. For each county, I calculated the amount of textile employment in the county and in neighboring counties within fifty miles. Denote this by $neighbor_c^{textile}$ for a given county c. The variable $neighbor_c^{textile}$ is above zero for 2,390 of the 3,140 counties in the United States. For those counties that have positive neighboring textile employment, I define the specialization fraction to be the fraction of the neighboring employment that is in specialized finishing plants,

$$Specialization\ Fraction_c \equiv \frac{neighbor_c^{finish}}{neighbor_c^{textile}}.$$

If there is greater vertical disintegration in agglomeration areas, the specialization fraction should be relatively high in counties with a high amount of neighboring textile employment.

Table 4 reports the mean and median specialization fractions across counties by neighboring employment size classes. The first group is counties with fewer than 100

employees in neighboring textile plants. Ignore this group for now. The next group is counties with between 100 and 499 in neighboring employees. The mean-specialization fraction across these counties is 0.057. The means for the next two size classes are approximately the same. The means of the last two size classes are substantially larger at 0.096 and 0.091. The table suggests that there is a critical agglomeration size of approximately 5,000 neighboring employees such that, when locations exceed this level, the mean specialization fraction increases from 0.05 to 0.09.

Now turn to the first group of counties with fewer than 100 in neighboring employment. The mean specialization fraction is 0.16, the highest of any group. This result might appear to be inconsistent with my hypothesis, but it actually has a simple explanation that is not damaging to my hypothesis. Suppose a mean fraction of 0.16 were to arise for this group on account of a typical county in this group having a single specialized finishing plant with 8 employees and another textile plant with 42 employees that is not a specialized finishing plant. If the latter plant were to weave unfinished cloth and then were to pass this along to the local finishing plant—and if this were the typical situation for counties in this group—it would certainly be inconsistent with my hypothesis that plants outside agglomerations tend to vertically integrated. However, this is not the typical situation with counties in this group. More than half of these counties have zero employment in specialized finishing plants. The high mean-specialization fraction is driven by the existence of counties that have a single plant that does only finishing (so the specialization fraction is 1 in this case). That is, there is not a division of labor in these counties where one plant specializes in weaving and another in finishing. Rather, for whatever reason, there happen to exist some counties that have isolated finishing plants that import unfinished textiles from outside the area.

This above discussion suggests the median might be a more useful statistic to look at to see what is happening in the typical county. For the median, there is no anomaly for the group with under 100 in employment. Examining the median reveals a clear pattern. The typical county with fewer than 5,000 in neighboring textile employment has a specialization fraction that is essentially zero. The typical county with 5,000–9,999 has specialization fraction of 0.06, while a county with more than 10,000 has a specialization fraction of 0.08. This result is consistent with my earlier finding that the purchased-inputs intensity measure is greater for plants inside agglomerations.

VI.   Conclusion

This paper considers purchased-inputs intensity (*PII*) as a measure of vertical disintegration. It finds that establishments located in areas where an industry is concentrated on average have a *PII* that is three percentage points higher than establishments located in areas where an industry is not concentrated. This three-percentage-point figure is arrived at through two different approaches. The first approach is simple; it considers only the most-concentrated industries and compares the *PII* of the leading state with the *PII* of the rest of the country. The second approach is somewhat complex. It develops procedures to estimate how *PII* varies with the amount of own-industry neighboring employment of a plant. The three-percentage-point difference cited is the average difference between plants with 10,000 to 25,000 in own-industry neighboring employment and plants with fewer than 500 in neighboring employment.

The paper estimates the average relationship for the entire manufacturing sector. In doing so, it ignores the fact that this relationship should be expected to vary across industries with different characteristics. For example, I have found that my estimated relationship is being driven by the industries that tend to be geographically concentrated. When I consider only concentrated industries (industries in the top half of the Ellison-Glaeser index), my estimate of the effect doubles. When I consider only industries in the bottom half, my estimate of the effect goes to zero. So industry characteristics matter. In future work, it would be useful to incorporate industry characteristics into the analysis.

REFERENCES

Adelman, M. A., "Concept and Statistical Measurement of Vertical Integration," in G. J. Stigler, (ed.), *Business Concentration and Price Policy* (Princeton, NJ: Princeton University Press, 1955), 281–322.
Allen, G. C., *The Industrial Development of Birmingham and the Black Country, 1860–1927* (London: Allen and Unwin, 1929).
Chinitz, B., "Contrasts in Agglomeration: New York and Pittsburgh," *American Economic Review* 51 (1961), 279–289.
Ciccone, Antonio, and Robert E. Hall, "Productivity and the Density of Economic Activity," *American Economic Review* 84 (1996), 54–70.
Ellison Glenn, and Edward Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* 105 (1997), 889–927.
Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer, "Growth in Cities," *Journal of Political Economy* 100 (1992), 1,126–1,152.
Hall, Max, *Made in New York: Case Studies in Metropolitan Manufacturing,* (Cambridge, MA: Harvard University Press, 1959).
Henderson, J. Vernon, "Efficiency of Resource Use and City Size," *Journal of Urban Economics* 19 (1986), 47–70.
Hoover, Edgar M., Jr., "Integration of Processes in Plant, Concern, and Production Center," in *Industrial Location and Natural Resources* (Washington, DC: United States Government Printing Office, 1943).
Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108 (1993), 577–598.
Krugman, Paul, "Increasing Returns and Economic Geography," *Journal of Political Economy* 99 (1991), 483–499.
Lichtenberg, Robert M., *One-Tenth of a Nation: National Forces in the Economic Growth of the New York Region,* (Cambridge, MA: Harvard University Press, 1960).
Marshall, Alfred, *Principles of Economics,* 8th edition (London: Macmillan, 1920).
Perry, Martin K., "Vertical Integration: Determinants and Effects," in Richard Schmalensee and Robert Willig (eds.), *Handbook of Industrial Organization,* Volume 1 (Amsterdam: North-Holland, 1989).
Rivera-Batiz, Francisco L., "Increasing Returns, Monopolistic Competition, and Agglomeration in Consumption and Production," *Regional Science and Urban Economics* 18 (1988), 125–153.
Rodríguez-Clare, Andrés, "Multinationals, Linkages, and Economic Development," *American Economic Review* 88 (1996), 852–873.
Rubenstein, James M., *The Changing U.S. Auto Industry: A Geographical Analysis* (London: Routledge, 1992).
Scott, Allen J., "Location and Linkage Systems: A Survey and Reassessment," *Annals of Regional Science* 17 (1983), 1–39.
Scott, Allen J., and E. C. Kwok, "Inter-Firm Subcontracting and Locational Agglomeration: A Case Study of the Printed Circuits Industry in Southern California," *Regional Studies,* 25 (1989), 405–416.
Scott, Allen J., and Doreen J. Mattingly, "The Aircraft and Parts Industry in Southern California: Continuity and Change from the Inter-War Years to the 1990s," *Economic Geography* 65 (1989), 48–71.
Stigler, George J., "The Division of Labor is Limited by the Extent of the Market," *Journal of Political Economy* 59 (1951), 185–193.
U.S. Bureau of Census, *Standard Industrial Classification Manual* (Washington, D.C.: Office of Management and Budget, 1987).

APPENDIX: THE REGRESSION PROCEDURE

This appendix presents a simple statistical model to motivate the empirical procedure that was employed. Consider a statistical process that operates at the level of each establishment. Suppose that the purchased-inputs intensity of establishment *e* in industry *i* in area *a* equals a linear function of the neighboring own-industry employment of the establishment plus a random variable,

$$\frac{purchased\ inputs_{eia}}{output_{eia}} = \alpha_i + \beta_i neighbor_{ia}^{own} + \epsilon_{eia} \qquad (5)$$

The first term, α_i, is a constant that may vary by industry. The second term is the product of a coefficient, β_i, that varies by industry and the level of own-industry neighboring employment of the establishment. The third term, ϵ_{eia}, is the difference between the realization of purchased-inputs intensity for the establishment and the conditional mean. Assume that ϵ_{eia} is i.i.d. throughout the universe of establishments, and suppose the variance is σ^2.

The average *PII* for establishments in industry/area *ia* equals the mean of the *PII* across establishments in the area with weights equal to the output shares of the establishments. Therefore, using equation (5), the average *PII* in industry/area *ia* can be written as

$$\frac{purchased\ inputs_{ia}}{output_{ia}} = \alpha_i + \beta_i neighbor_{ia}^{own} + \epsilon_{ia}, \qquad (6)$$

where ϵ_{ia} is the weighted sum of the establishment disturbances in the industry/area,

$$\epsilon_{ia} \equiv \sum_{e \in ia} w_{eia}\epsilon_{eia}, \qquad (7)$$

with the weights equal to the output shares w_{eia} of the establishments, defined in the text by equation (2). The variance of ϵ_{ia} relative to the variance of the establishment-level disturbances equals

$$\frac{\sigma_{ia}^2}{\sigma^2} = \sum_{e \in ia} w_{eia}^2. \qquad (8)$$

With the industry/area data set, the parameters α_i and β_i can be estimated using weighted least squares with each observation weighted by the inverse of the relative variance (8). As discussed in the text, the actual output share w_{eia} is not observed, so I instead used the estimated employment share $\hat{w}_{eia}$ defined by equation (3). I then plug these estimated

employment shares into equation (8) to obtain an estimate of the relative variance and use the inverse as the weight in the weighted-least squares procedure.

Consider the hypothesis that all the industries in a certain group of industries have the same slope term but can vary in the intercept term; i.e., $\beta_i = \beta$ for all industries i in this group, but $\alpha_i \neq \alpha_{i'}$ can happen for $i \neq i'$. In this case, the parameter β can be estimated by differencing the data by the industry means of each variable using as weights the inverse of the relative variance (8). I differenced the data in this way and used the differenced data for all of the analysis. This procedure removes the industry fixed effect α_i from the data. Because of this differencing of the data, no constant terms are reported.