

The Determinants of Agglomeration¹

Stuart S. Rosenthal

*Department of Economics, Center for Policy Research, Syracuse University, Syracuse,
New York 13244-1020*

E-mail: ssrosent@maxwell.syr.edu

and

William C. Strange

*Faculty of Commerce and Business Administration, University of British Columbia,
Vancouver, British Columbia V6T 1Z2, Canada*

E-mail: william.strange@commerce.ubc.ca

Received May 23, 2000; revised April 23, 2001; published online July 20, 2001

This paper examines the microfoundations of agglomeration economies for U.S. manufacturing industries. Using industries as observations, we regress the Ellison–Glaeser (G. Ellison and E. Glaeser, 1997, *J. Polit. Econ.* **105**, 889–927) measure of spatial concentration on industry characteristics that proxy for the presence of knowledge spillovers, labor market pooling, input sharing, product shipping costs, and natural advantage. The analysis is conducted separately at the zipcode, county, and state levels. Results indicate that proxies for labor market pooling have the most robust effect, positively influencing agglomeration at all levels of geography. Proxies for knowledge spillovers, in contrast, positively affect agglomeration only at the zipcode level. Reliance on manufactured inputs or natural resources positively affects agglomeration at the state level but has little effect on agglomeration at lower levels of geography. The same is true for the perishability of output, a proxy for product shipping costs. © 2001 Academic Press

Press

1. INTRODUCTION

A growing empirical literature has established that the spatial concentration of manufacturing activity enhances productivity and growth (e.g., Moomaw

An earlier version of this article was published on IDEAL *First* (www.idealibrary.com) but was retracted because it contained inaccurate estimates caused by a programming error. The article published here contains the corrected estimates.

¹ We gratefully acknowledge the financial support of the Social Sciences and Humanities Research Council of Canada, the UBC Centre for Real Estate and Urban Land Economics, the Real Estate Foundation of British Columbia, and the Center for Policy Research at Syracuse University. We thank David Audretsch for providing us with data on innovations and Jan Brueckner, Vernon Henderson, and two anonymous referees for helpful comments. We also thank Peter Howe for valuable research assistance and Esther Gray for help in preparing the manuscript.

[18], Sveikauskas [23], Nakamura [19], Henderson [12], and Ciccone and Hall [3]). These studies show that localization economies—economies of scale arising from spatial concentration of activity within industries—are of particular importance. Urbanization economies—economies of scale arising from city size itself—although important, have smaller effects on productivity. Glaeser, Kallal, Scheinkman, and Shleifer [9] and Henderson, Kuncoro, and Turner [13] demonstrate the importance of these sorts of increasing returns. Ellison and Glaeser [5] establish that the level of agglomeration varies considerably across industries, as does the tendency of an industry to coagglomerate with other industries.

This literature has had relatively little to say about the causes of agglomeration. Two notable exceptions are Audretsch and Feldman [1] and Dumais, Ellison, and Glaeser [4]. Audretsch and Feldman use a spatial Gini coefficient to measure geographic concentration. They show that innovative activity is substantially more concentrated than overall production and that industries that emphasize research and development tend to be more spatially concentrated.² Dumais *et al.* [4] look at the microfoundations of agglomeration economies by considering which industries coagglomerate. They find that industries with similar labor mixes enjoy the largest benefit from proximity, suggesting the importance of labor market pooling.

In contrast, theoretical work in this area has had much more to say about the causes of agglomeration. It has been demonstrated that agglomeration economies can arise from labor market pooling (Helsley and Strange [11]), input sharing (Goldstein and Gronberg [10]), and knowledge spillovers (Glaeser [7]). See Quigley [21] for a survey of the extensive theoretical literature on the microfoundations of agglomeration economies.

This paper bridges the empirical and theoretical literatures. Utilizing the Ellison and Glaeser [5] index, we measure the level of spatial concentration among manufacturing industries at the zipcode, county, and state levels in the fourth quarter of 2000. The Ellison–Glaeser index depends on both the geographic distribution of employment and the intraindustry allocation of employment to establishments. The expected value of the Ellison–Glaeser index is zero when the spatial allocation of employment is random. Thus, the index has the appealing feature of permitting comparisons between the actual pattern of spatial concentration and the concentration that would be expected to arise from a random allocation of employment.³

Matching geographic concentration measures with data on industry characteristics, we seek to explain differences in the spatial concentration of industries.

² A related result is obtained by Jaffe *et al.* [14], who identify a “paper trail” of knowledge spillovers in the location of patent citations. They show that patent citations are highly spatially concentrated, with citations 5 to 10 times as likely to come from the same SMSA as control patents.

³ As will become apparent, the Ellison–Glaeser index converges to the more widely known spatial Gini measure of agglomeration as the industry structure approaches that of a perfectly competitive market.

We focus on the three microfoundations of agglomeration that have been most prominent in the theoretical literature: knowledge spillovers, labor market pooling, and input sharing. Our approach is to regress the Ellison–Glaeser localization index on industry characteristics that proxy for the three causes of localization and on controls for product shipping costs and natural advantage. The regressions are carried out using 4-digit manufacturing industries as observations. All of the regressions are carried out separately for concentration at the zipcode, county, and state levels, since the causes of agglomeration could well differ at different levels of geographic aggregation.

Results provide evidence of the importance of all three sources of localization economies. The evidence is strongest for labor market pooling, with proxies having a positive impact on agglomeration at all levels of geography. The proxies for knowledge spillovers impact agglomeration positively only at the zipcode level. Reliance on manufactured inputs or natural resources positively affects agglomeration at the state level but has little effect on agglomeration at lower levels of geography. The same is true for the perishability of output, a proxy for product shipping costs. In contrast, reliance on service inputs reduces state-level agglomeration. Taking all of these results together, an interesting pattern emerges, with shipping-oriented attributes (manufactured inputs, resources, perishability) influencing agglomeration at the state level, knowledge spillovers impacting highly localized agglomeration, and labor impacting agglomeration at all levels of geography. These findings are largely robust, holding for both ordinary least squares (OLS) and 2-digit standard industry classification (SIC) fixed-effect specifications, as well as for alternative Metropolitan Statistical Area (MSA) based measures of geography, and when industries are aggregated from the 4-digit to the 3-digit level.

The remainder of the paper is organized as follows. Section 2 considers the degree to which industries agglomerate. Section 3 looks at the determinants of an industry's agglomeration. Section 4 concludes the paper.

2. THE EXTENT OF AGGLOMERATION

2.1. *An Index of Agglomeration*

This section addresses the degree to which industries agglomerate. There are a number of statistics that one might employ to characterize the degree of agglomeration. A natural candidate is the spatial Gini coefficient, defined as $G \equiv \sum_i (x_i - s_i)^2$, where x_i is location i 's share of total employment and s_i is the location's share of employment in a particular industry. This statistic is employed by Krugman [16] and Audretsch and Feldman [1], among others. It takes on a value of zero when an industry is allocated across space in exactly the same way as for total employment. It takes on a value close to one (depending on the size of the industry itself) when the industry is completely concentrated in one location.

As Ellison and Glaeser [5] note, however, $G > 0$ does not necessarily imply that the industry in question is overconcentrated. Suppose that an industry is made up of a small number of large plants and that there is no agglomerative force—either an externality or a natural advantage—leading to concentration. In this case, G will take on a large value simply because of the industrial organization of the industry. In Ellison and Glaeser's metaphor, tossing three darts will leave most of the dartboard without any darts. The spatial Gini coefficient, therefore, does not distinguish random concentration arising from industrial structure from concentration arising from agglomerative externalities or natural advantage.

To address this problem, Ellison and Glaeser propose the following index of concentration:

$$\gamma = \frac{G - (1 - \sum_i x_i^2)H}{(1 - \sum_i x_i^2)(1 - H)}. \quad (2.1)$$

$H = \sum_j z_j^2$ is a Herfindahl index of the J plants in the industry, with z_j representing the employment share of the j th plant. For a perfectly competitive industry with a large number of small plants, H approaches zero and γ approaches $G/(1 - \sum_i x_i^2)$.⁴ In this case, G measures spatial concentration without any contamination associated with industrial organization. More generally, γ takes on a value of zero when an industry is as concentrated as one would expect from a random location process, while a positive value of γ indicates excess concentration. As Ellison and Glaeser take pains to point out, however, a positive γ does not necessarily indicate that agglomerative externalities are present. Instead, agglomerative externalities and natural advantage are in a sense observationally equivalent. Observing that the industry is concentrated does not identify the cause of the concentration.

2.2. Which Industries Agglomerate?

We compute the Ellison–Glaeser index using information from Dun and Bradstreet (D & B), included in the IMarket Inc. MarketPlace database for the fourth quarter of 2000.⁵ The complete version of the data set contains establish-

⁴ The $(1 - \sum_i x_i^2)$ term is included in order that the index have the property that $E(\gamma) = 0$ when neither agglomerative spillovers nor natural advantage are present (see Ellison and Glaeser [5] for details). For the state, county, and zipcode levels that we consider, $(1 - \sum_i x_i^2)$ is close to one, taking on values of 0.9997, 0.9954, and 0.9578, respectively.

⁵ IMarket Inc. is a commercial data vendor. IMarket obtains the core data in the MarketPlace file from Dun and Bradstreet, another commercial data vendor, then matches the D & B data with a wide variety of other data from other data vendors, and packages all of these data together in the MarketPlace file. The analysis in this paper is based solely on the D & B portion of the MarketPlace database. In addition, although earlier versions of this paper were based on data from 1996, we focus here on data from the fourth quarter of 2000. This is because representatives at IMarket Inc. advised us that the more recent data is of higher quality and somewhat more complete.

ment-level information on over 12 million establishments in the United States. We utilized a more manageable and affordable version of the data set in which the data were aggregated up to the zipcode level.⁶ In phone conversations with analysts at D&B, we were advised that firms requesting not to be in the database are omitted from the data file. Partly for that reason, the D&B database, while extensive, does not contain the entire universe of establishments in the United States. Nevertheless, the D&B analysts felt that the omissions from the data set are sufficiently random that the D&B database is representative of the spatial distribution of establishments in the United States.⁷

We calculate the location statistic γ at the state, county, and zipcode levels separately for manufacturing industries using three different definitions of industries based on 2-, 3-, and 4-digit SIC codes. As is apparent in Table 1a, for each level of geography, the average level of agglomeration increases as one goes from 2- to 3-digit industries and from 3- to 4-digit industries. This occurs because as industries become aggregated into ever broader and fewer categories, spatial patterns of establishment locations eventually approach that of the entire economy, causing G and γ to shrink toward zero.⁸ For this reason, the remainder of our discussion is based primarily on 4-digit-level industries, though we will on occasion examine features of 2- and 3-digit-level industries for comparison.

Focusing on the 459 4-digit manufacturing industries, at the state level the mean level of agglomeration, $\bar{\gamma}_s$, is 0.0485.⁹ At the county and zipcode levels the means are $\bar{\gamma}_c = 0.0193$ and $\bar{\gamma}_z = 0.0101$, respectively. Since γ equals zero when an industry is as concentrated as a random allocation, whenever $\gamma > 0$ there is excess concentration while $\gamma < 0$ implies an excess diffusion of

⁶ Additional details on the Dun and Bradstreet (D&B) MarketPlace file are provided at the Dun and Bradstreet web site, www.dnb.com. As described by Dun and Bradstreet, there are several important benefits to firms from listing themselves in the D&B database and obtaining a D-U-N-S identification number. These benefits arise primarily because of the incredible size of the D&B data file. Because the D&B file is such an effective source of information on firms throughout the economy, businesses use the D&B file to do market analysis and search out potential trading partners. Individual firms therefore have an incentive to list themselves with D&B in much the way firms have an incentive to voluntarily list themselves in the yellow pages. In addition, DUNS identification numbers are rapidly becoming a standard identification device in the economy, and many companies including the Federal Government require that clients obtain a D-U-N-S number as a precondition for engaging in trade. As noted in the D&B website, "It [the D-U-N-S number] is now the standard for all United States Federal Government electronic commerce transactions to help streamline and reduce federal procurement costs."

⁷ In contrast, the Census of Manufacturing (CM) and County Business Patterns (CBP), the data sets used by Ellison and Glaeser (1997), are designed as representative surveys. However, the CM and CBP both suffer from restrictions on the type of firms and employment data reported, including top-coding problems. There is no top-coding in the IMarket database.

⁸ In the limit, with a single industry category, industry employment is coincident with the entire manufacturing sector, G equals zero, H approaches zero, and γ goes to zero.

⁹ See Table 1a for additional summary statistics.

TABLE 1a
 Summary Measures of Agglomeration among Manufacturing Industries at the
 SIC 2-Digit, 3-Digit, and 4-Digit Levels

γ	Mean	SD	Min	Max	Correlation with gamma at the	
					State level	County level
2-Digit						
State	0.0241	0.0449	0.0017	0.1946	1.00	
County	0.0099	0.0284	0.0010	0.1300	0.91	1.00
Zip code	0.0029	0.0088	0.0002	0.0399	0.91	1.00
3-Digit						
State	0.0433	0.0706	-0.0024	0.4078	1.00	
County	0.0163	0.0416	-0.0002	0.3718	0.78	1.00
Zip code	0.0082	0.0342	-0.0009	0.3515	0.70	0.87
4-Digit						
State	0.0485	0.0717	-0.0709	0.4993	1.00	
County	0.0193	0.0379	-0.0131	0.3718	0.82	1.00
Zip code	0.0101	0.0275	-0.0046	0.3515	0.58	0.73

employment. As noted by Ellison and Glaeser [5], it is not obvious how to decide what levels of γ constitute significant departures from a random allocation. Looking at values of state-level γ for notably concentrated industries like computers and automobiles, they define $\gamma > 0.05$ as highly concentrated and $\gamma < 0.02$ as not very concentrated. Applying these benchmarks to the average values of γ for our data, it is apparent that there is clear evidence of excess concentration at the state level. Moreover, only 18 of the 459 state-level γ values are below zero (implying excess dispersion in those industries). At the county and zipcode levels, the average level of agglomeration across the industries is much lower, but once again only a small number of industries have negative γ values (9 industries at the county level and 9 at the zipcode level). Finally, observe that the correlation between the state- and county-level γ among 4-digit industries is 82% while the correlation between state- and zipcode-level γ is only 58%. Together, these results and those above suggest that the process generating state-level agglomeration is different than the one generating agglomeration at the county and zipcode levels, a theme that will recur at various points in the discussion to follow.

Because Ellison and Glaeser [5] examined agglomeration only down to the county level, our measures of zipcode-level concentration are new to the literature, and some discussion of the pattern of agglomeration at that level is warranted, especially for those industries whose agglomeration has become part of the geographic folklore. One such industry is the carpet industry, SIC 2273, the history of whose localization was discussed by Krugman [16]. This industry

shows a considerable degree of agglomeration at every level of geography, with $\gamma_s = 0.406$, $\gamma_c = 0.089$, and $\gamma_z = 0.048$. The motor vehicle industry is denoted SIC 3711. It has $\gamma_s = 0.089$, $\gamma_c = 0.020$, and $\gamma_z = 0.0027$. In the case of motor vehicles, there is excess agglomeration at the state level, but much less agglomeration at the county level and little excess agglomeration at the zipcode level. Based on these comparisons and the results in Table 1a, it is apparent that there is always less agglomeration at the zipcode level, and notoriously agglomerated industries may not even appear to be agglomerated at a microgeographic level, at least relative to a random allocation of employment across space.

Table 1b lists the level of agglomeration at the state, county, and zipcode levels for all twenty 2-digit industries. Although Tobacco (SIC 21) and Textiles (SIC 22) display considerable excess agglomeration, especially at the state level, most of the industries display relatively little excess agglomeration, although all of the γ values are positive. Of course, as noted above, the high degree of industry aggregation at the 2-digit level obscures much of the variation in spatial concentration across industries.

TABLE 1b
Agglomeration of Manufacturing Industries at the SIC 2-Digit Level

SIC	Definition	State γ	County γ	Zip code γ
20	Food and kindred products	0.00347	0.00119	0.00029
21	Tobacco manufactures	0.19457	0.13002	0.03989
22	Textile mill products	0.09410	0.00601	0.00177
23	Apparel and related products	0.01159	0.00653	0.00184
24	Lumber and wood products, except furniture	0.01168	0.00284	0.00034
25	Furniture and fixtures	0.01212	0.00297	0.00074
26	Paper and allied products	0.00844	0.00213	0.00035
27	Printing, publishing, and allied products	0.00527	0.00264	0.00039
28	Chemicals and allied products	0.01047	0.00369	0.00062
29	Petroleum refining and related products	0.03605	0.01040	0.00428
30	Rubber and miscellaneous plastics products	0.00385	0.00102	0.00023
31	Leather and leather products	0.01513	0.00640	0.00298
32	Stone, clay, glass, and concrete products	0.00357	0.00209	0.00052
33	Primary metal products	0.01438	0.00202	0.00041
34	Fabricated metal products, except machinery and transportation equipment	0.00447	0.00095	0.00021
35	Machinery, except electrical	0.00170	0.00112	0.00029
36	Electrical and electronic machinery, equipment, and supplies	0.00869	0.00352	0.00050
37	Transportation equipment	0.02203	0.00462	0.00084
38	Scientific and professional instruments; photographic and optical goods; watches	0.01453	0.00429	0.00018
39	Miscellaneous manufactured commodities	0.00666	0.00306	0.00055

Table 1c provides a sharper picture by listing the 10 most concentrated industries at the state, county, and zipcode levels for the 4-digit industries. This table has a number of interesting implications. First, some of the most agglomerated industries may well be agglomerated because of natural advantage rather than because of a spatial externality. Fur goods (SIC 2371) and Cigarettes (SIC 2111) are examples of this. Second, many of the agglomerated industries are the kinds of manufacturing industries where one might expect agglomeration economies to be important. Guided missiles & space vehicles (SIC 3761) and Office machines (SIC 3579) are examples of this. Third, although there are some industries that are highly agglomerated at more than one level of geography, for the most part the lists are distinct. This finding provides further support for the idea that different processes generate agglomeration at different levels of geography.

3. THE DETERMINANTS OF AGGLOMERATION

3.1. *Overview*

The central goal of this section is to evaluate the degree to which agglomerative externalities explain interindustry differences in spatial concentration. Accordingly, our strategy is to regress γ on proxies for three key sources of agglomerative spillovers: knowledge spillovers, labor market pooling, and input sharing. We also provide controls for natural advantages and product shipping costs. Summary statistics and data sources are provided in Table 2a at the 4-digit level for the manufacturing sector.

3.2. *Controls for Natural Advantage and Transportation Costs*

It has long been recognized that natural advantages can affect the location decisions of firms because of both the cost of shipping inputs to the factory and the cost of shipping output to the market. From that observation, it is a short step to recognizing that natural advantage can also influence an industry's spatial concentration. Kim [15] estimates a state-level Rybczynski equation relating employment to factor endowments, assuming that all factors of production are immobile, including labor. He argues that the residuals in this estimation are upper bounds on the strength of agglomeration economies. In a similar way, Ellison and Glaeser [6] employ predicted state-level employment variables to account for the importance of natural advantage in agglomeration. Both Kim [15] and Ellison and Glaeser [6] conclude that natural advantage is important.

We use several variables from the 1992 Bureau of Economic Analysis (BEA) input-output tables to control for the importance of natural advantages associated with proximity to inputs. The variables *Energy per \$ shipment*, *Natural resources per \$ shipment*, and *Water per \$ shipment* measure energy input cost, the cost of natural resources, and water-related costs respectively as

TABLE 1c
The 10 Most Agglomerated Manufacturing Industries at the SIC 4-Digit Level

SIC	Zipcode level		County level		State level		
	SIC description	γ	SIC	SIC description	SIC	SIC description	γ
2371	Fur goods	0.352	2371	Fur goods	2397	Schiffli machine embroideries	0.499
3761	Guided missiles & space vehicles, parts	0.260	2397	Schiffli machine embroideries	3761	Guided missiles & space vehicles, parts	0.434
3579	Office machines & parts	0.145	3761	Guided missiles & space vehicles, parts	2284	Thread and handwork yarns	0.413
2087	Flavoring extracts & syrups	0.142	2874	Phosphatic fertilizers	2371	Fur goods	0.408
3149	Footwear, except rubber, n.e.c.	0.139	3861	Photographic equipment and supplies	2273	Carpets and rugs	0.406
2335	Womens' and misses' dresses	0.118	2111	Cigarettes	2084	Wines, brandy, and brandy spirits	0.372
2381	Fabric dress and work gloves	0.114	3149	Footwear, except rubber, n.e.c.	2251	Womens' hosiery, except socks	0.371
3764	Missile and rocket engines	0.111	2043	Cereal breakfast foods	3533	Oil & gas field equipment & parts	0.339
3676	Electronic resistors	0.086	2335	Women's, misses', and juniors'	2436	Softwood veneer and plywood	0.328
3844	X-ray apparatus, tubes, & parts	0.084	2841	Soap & detergents	2141	Manufactured tobacco	0.305

TABLE 2a
4-Digit Manufacturing Industries: Definitions, Data Sources, and Selected Summary Statistics for Explanatory Variables

Variable name	Definition	Data source and time period	No. of obs.	Mean	SD	Min	Max
Innovations from firms with fewer than 500 workers per \$ shipment	Number of new products in 1982 trade magazines for firms with < 500 employees divided by dollar value of shipments (in \$1,000,000)	1982 U.S. SBA Innovation Data Base and 1992 Annual Survey of Manufacturers from NBER	459	7.22×10^{-4}	1.72×10^{-3}	0.0000	0.0146
Innovations from firms with more than 500 workers per \$ shipment	Number of new products in 1982 trade magazines for firms with > 500 employees divided by dollar value of shipments (in \$1,000,000)	1982 U.S. SBA Innovation Data Base and 1992 Annual Survey of Manufacturers from NBER	459	7.55×10^{-4}	1.84×10^{-3}	0.0000	0.0188
Shipments net of inputs per worker ^a	Shipments - Material costs (in \$1,000,000) divided by total employees (in 1,000)	1992 Annual Survey of Manufacturers from NBER	459	86.2	70.8	24.0	975.6
Managerial share of workers ^a	(Total employees - Production employees) divided by total employees	1992 Annual Survey of Manufacturers from NBER	459	0.2869	0.1219	0.0781	0.8270
Share of workers with Ph.D. or professional degree ^a	Share of employed individuals with Ph.D. or professional degree	1995 Consumer Population Survey March file	427	0.0092	0.0203	0.0000	0.1765
Share of workers with Master's degree ^a	Share of employed individuals with Master's degree	from Unicon Inc. 1995 Consumer Population Survey March file	427	0.0332	0.0342	0.0000	0.1637
Share of workers with Bachelor's degree ^a	Share of employed individuals with Bachelor's degree	from Unicon Inc. 1995 Consumer Population Survey March file	427	0.1198	0.0686	0.0000	0.3455

Manufactured inputs per \$ shipment	Cost of inputs obtained from manufacturing firms (SIC 20 through 39) per dollar shipments	459	0.3502	0.1166	0.0063	0.6572
Nonmanufactured inputs per \$ shipment ^b	Cost of materials other than manufactured inputs, energy, natural resources, and water per dollar of shipments	459	0.0819	0.0675	-0.1119	0.6557
Natural resources expenses per \$ shipment	Cost of natural resource inputs per dollar of shipments	459	0.0357	0.1026	0.000	0.7923
Energy expenses per \$ shipment	Cost of energy inputs per dollar of shipments	459	0.0208	0.0232	0.0022	0.2405
Water expenses per \$ shipment	Cost of water inputs per dollar of shipments	459	0.0018	0.0029	0.0000	0.0238
Inventories per \$ shipment (nonperishability) ^c	Dollar value of end-of-year inventories per dollar of shipments	459	0.1446	0.0634	0.0209	0.5053

^aSee Table 2b for additional details.

^bSee Appendix Table A-2 for additional details.

^cSee Table 2c for additional details.

fractions of the value of shipments. These variables were available at the 4-digit level.¹⁰ To the extent that industries concentrate because of a desire to locate close to the sources of their energy, natural resource, and water related inputs, we expect the coefficients on these variables to be positive.¹¹

It has also long been recognized that the cost of transporting output can affect location decisions. A tempting approach to control for such effects would be to use readily available BEA data on actual product shipping costs by industry. This, however, would not be suitable because industries for which the per mile cost of shipping the product is high would locate so as to minimize distances to their markets and the related shipping costs. Instead, we proxy for the per mile cost of shipping the product using *Inventories per \$ of shipment*, defined as the value of end-of-year inventories divided by the value of shipments. Industries that produce highly perishable products face high product shipping costs per unit distance and, therefore, will seek to locate close to their markets, *ceteris paribus*. With multiple markets, such industries will tend to display less agglomeration. Conversely, industries that produce nonperishable products face lower product shipping costs and should display more agglomeration.¹²

Table 2b provides compelling support for using *Inventories* to proxy for perishability. The table displays the ten 4-digit industries with the highest values of *Inventories* and the ten industries with the lowest values of *Inventories*. Industries with very low inventory–shipment ratios include meat packing plants, newspapers, milk and cream, and other clearly perishable products. Industries with the highest inventory–shipment ratios include aircraft, wine and other liquors, machinery, and other clearly nonperishable products. These data on *Inventories* were obtained from the 1992 Annual Survey of Manufactures which was obtained at the NBER website (www.nber.org). To the extent that industries concentrate when per-mile costs of shipping the product are low, we expect the coefficients on this variable to be positive.

¹⁰ The URL for the 1992 BEA Input–Output file is <http://www.bea.doc.gov/bea/dn2/i-o.htm>. The file is zipped and downloadable. The file name is “1992 Benchmark I–O Table Six-Digit Transactions” and contains the make table, use table, direct requirements coefficients table, and estimates by commodity of transportation costs and of wholesale and retail margins (498-industry detail). Once unzipped there are a number of files, including instructions on how to make an extract from the data sets. In addition, the input–output tables are organized by product type rather than by SIC category. We obtained a concordance from BEA to match the product types to 4-digit SIC categories.

¹¹ A detailed description of the SIC categories used to construct the category *Natural Resources* is provided in the appendix. Note that coal, crude petroleum, and natural gas are included in the *Energy* variable rather than in *Natural resources*.

¹² Of course, other factors besides perishability of the product affect optimal inventory–shipment ratios. For example, internal economies of scale create incentives for firms to produce in bulk and stockpile output for later shipment. It is worth pointing out that internal economies of scale also directly influence agglomeration through their impact on the size distribution of establishments. However, that is already dealt with through the inclusion of the Herfindahl index in γ .

TABLE 2b
Industries with the Lowest and Highest Inventory-to-Shipment Ratios

Lowest inventory-to-shipment ratio			Highest inventory-to-shipment ratio		
SIC	SIC description	Inv/ship	SIC	SIC description	Inv/ship
2011	Meat packing plants	0.021	3721	Aircraft	0.505
2813	Industrial gases	0.021	2084	Wines, brandy, and brandy spirits	0.470
2711	Newspapers	0.022	2085	Distilled and blended liquors	0.347
2026	Fluid milk and cream, and related products	0.023	3541	Machine tools, metal cutting types and parts	0.342
2051	Bread and other bakery products	0.025	3533	Oil and gas field equipment and parts	0.342
2096	Potato chips, corn chips, and similar products	0.027	3262	Vitreous china table and kitchenware	0.323
3711	Motor vehicles and passenger car bodies	0.027	2063	Beet sugar	0.314
2021	Creamery butter	0.032	3542	Machine tools, metal forming types	0.293
2015	Poultry slaughtering and processing	0.035	3511	Steam, gas, and hydraulic turbines	0.291
2082	Malt beverages	0.035	3356	Extruded nonferrous metal mill products	0.291

3.3. Controls for Agglomerative Externalities

Two variables are used to proxy for input sharing. *Manufactured inputs per \$ of shipment* is the ratio of the cost of inputs purchased from the manufacturing sector—SIC codes 20 to 39—to the value of shipments. This variable was obtained from the 1992 BEA input–output tables and measures the relative importance of manufactured inputs for the industry. Among industries for which *Manufactured inputs* is large, the gains from sharing inputs are likely to also be large, creating incentives to concentrate spatially. For that reason, we expect *Manufactured inputs* to have a positive coefficient. Similarly, we also include a variable *Nonmanufactured inputs per \$ of shipment*, where *Nonmanufactured inputs* is the value of materials other than those already noted (manufactured inputs, energy, natural resources, and water).¹³ This category of inputs includes such things as legal services, accounting and financial services, insurance, communication, repair, and janitorial services. There are two impor-

¹³ *Nonmanufactured inputs* is measured as a residual and is calculated by subtracting our other input measures and value added per dollar of shipments from unity since shipments are approximately equal to value added plus expenditures on materials. A detailed list of the SIC categories that comprise the *Nonmanufactured inputs* is provided in the appendix. In addition, data on the value added and shipments used to construct *Nonmanufactured inputs* was obtained from the 1992 Annual Survey of Manufactures while the other variables used to construct *Nonmanufactured inputs* were obtained from the 1992 BEA input–output tables as noted above.

tant differences between manufactured and nonmanufactured inputs. First, scale economies are likely to be stronger for manufactured inputs. Second, manufactured inputs are likely to exhibit greater industry specificity. For both of these reasons, there is less reason for industries that rely heavily on nonmanufactured inputs to agglomerate. Accordingly, we expect *Nonmanufactured inputs* to have less impact on agglomeration than does *Manufactured inputs*.

The variable used to proxy for the importance of knowledge spillovers is *Innovations per \$ of shipment*. Innovations are defined as the number of new products advertised in trade magazines in 1982, the only year for which such data were readily available. An essential input for innovation is new knowledge. In that regard, innovative activity is related to the importance of knowledge spillovers. In addition, although our innovation variable predates our agglomeration measures by 18 years, it seems likely that most industries for which innovation was important in 1982 would continue to place importance on innovation in the 1990s. Accordingly, we anticipate that *Innovations per \$ of shipment* will have a positive effect on our industry concentration measures. The innovation data were collected by the U.S. Small Business Administration as part of its Innovation Database and were available at the 4-digit level. See Audretsch and Feldman [1] for additional details on these data.¹⁴

There is reason to believe that the operation of knowledge spillovers is linked to the industrial organization of an industry. Saxenian [22], for instance, argues that the open managerial structure of the high-technology firms in Silicon Valley gave it an advantage over the relatively closed structure typical of the large high-tech firms populating Boston's Route 128. Consistent with that argument, Rosenthal and Strange [20] find that smaller establishments have a larger effect on the attractiveness of a location than do larger establishments, *ceteris paribus*. In addition, Audretsch, van Leeuwen, Menkveld, and Thurik [2] find that small establishments are more productive than large establishments, *ceteris paribus*. To allow for the possibility that innovativeness has different effects on agglomeration depending on the size of the firms that innovate, we partition the *Innovations* variable into innovations at firms with fewer than 500 employees and innovations at firms with more than 500 employees.¹⁵

¹⁴ We are grateful to David Duretsch for providing these data.

¹⁵ Two other variables were considered but rejected as proxies for the importance of knowledge spillovers. The first is the number of patents. However, patents are not really the same as innovations. In some industries, a single innovation can be associated with hundreds of patents. In addition, the U.S. Patent Office codes patents based on the product type, not the industry to which the innovating firm belongs. Thus, it is difficult to accurately match patent data to the SIC definitions of industries. Another candidate variable as a proxy for the importance of information spillovers would be industry expenditures on research and development. However, because many innovations are associated with business practice rather than the deliberate search for new products or processes, this variable does not provide as precise a measure of the importance of information spillovers as do the innovations. In addition, expenditures on research and development are indirectly related to the role of information spillovers in that they are an input rather than an output.

The most difficult of the Marshallian microfoundations to proxy is labor market pooling. If pooling is possible, an industry benefits by agglomerating because it is better able to hire workers with industry-specific skills. The problem in proxying for the importance of pooling in an industry is that it is difficult to identify industry characteristics that are related to the specialization of the industry's labor force. We therefore separately employ three different proxies. The first is *Net productivity*, equal to the value of shipments less the value of purchased inputs, all divided by the number of workers in the industry. This measure of the productivity of labor is obtained by using the ASM data for 1992 taken from the NBER website as described above. The second is the ratio *Management workers/(Management + Production workers)*. This "brains to brawn" variable measures the share of supervisory and support labor in production. If little of such labor is needed, then production is more likely to be a matter of routine, and specialized labor is likely to be less important. This variable is also constructed using data from the 1992 ASM. The final approach to proxying for labor market pooling is to employ variables on worker education, specifically the percentage of workers with *Doctorates*, *Master's Degrees*, and *Bachelor's Degrees*.¹⁶ These data are obtained from Consumer Population Survey (CPS) data from 1995.¹⁷

It is worth noting that all of these proxies for the importance of labor market pooling are positively correlated, as shown in Table 2c. For example, correlation between *shipments net of inputs per worker* and the other proxies for labor market pooling range between 21 and 31%. Correlation between *Managerial share of workers* and *Share of workers with Master's degrees* is 53%. Given the strong positive correlation between these variables, the models to follow are all estimated separately for each of the three sets of labor market pooling proxies. In all cases these variables are expected to have positive coefficients.

¹⁶ It is important to note that while educated workers may indeed be specialized, these variables do not capture the degree to which less-educated workers may also have specialized industry-specific skills (i.e., Marshall's [17] cutlery manufacturers).

¹⁷ The CPS reports the industry of occupation for individual workers. We computed the distribution of employed workers across such industry categories, and then matched industry codes to SIC categories using a correspondence table provided at the census website (www.bls.census.gov/cps/bindcd.htm). It is worth noting that CPS industry codes correspond to 3-digit SIC codes with the exception of two industry codes that match to 2-digit SIC codes, and one industry code that matches directly to a 4-digit SIC code. In order to use these data for 4-digit-level analysis, therefore, we assigned the 3-digit SIC education values to 4-digit member subgroups in the SIC classification scheme. Unfortunately, this precludes using the education variables when 3-digit SIC fixed effects are included in some of the models since the education variables do not vary within 3-digit SIC classifications.

TABLE 2c
Correlation Between Proxies for the Importance of Labor-Market Pooling^a

	Shipments net of inputs per worker	Managerial share of workers	Share of workers with Ph.D. or professional degree	Share of workers with Master's degree	Share of workers with Bachelor's degree
Shipments net of inputs per worker	1.00				
Managerial share of workers	0.27	1.00			
Share of workers with Ph.D. or professional degree	0.24	0.28	1.00		
Share of workers with Master's degree	0.31	0.53	0.50	1.00	
Share of workers with Bachelor's degree	0.21	0.57	0.45	0.56	1.00

^aSample size equals 427 4-digit industries.

3.4. *Estimates of the Determinants of Agglomeration*

The effect of agglomerative spillovers on the spatial concentration indexes is measured by estimating

$$\gamma_{j,m} = \beta X_m + \varepsilon_{j,m}, \quad (3.1)$$

where $\gamma_{j,m}$ is the localization statistic for the m th industry at level of geography j , X_m is the vector of industry characteristics with associated coefficient vector β , and $\varepsilon_{j,m}$ is assumed to be an independent and identically distributed error term. We estimate Eq. (3.1) separately for the three geographic specifications, with γ_j measured at the state, county, and zipcode levels.

Before proceeding further, it is important to discuss identification. Because the role of natural advantages and product shipping costs in an industry is likely to be exogenous to the level of agglomeration, coefficient estimates on these variables provide direct measures of their impact on concentration. For the remaining variables, the coefficients describe the equilibrium relationship between industry characteristics and agglomeration: industry characteristics affect the propensity to agglomerate, but agglomeration can influence industry characteristics. In both directions, however, these relationships are governed by the degree to which agglomeration reduces costs. Specifically, agglomeration reduces the cost of innovation by enhancing knowledge spillovers while also

reducing the cost of labor and intermediate inputs through labor market pooling and input sharing. Precisely for these reasons, industries sensitive to innovation, labor, and intermediate input costs are more likely to agglomerate. Thus, evidence of a positive relationship between agglomeration and these other factors confirms that tendencies to innovate, pool labor, and share inputs all lead to an increase in agglomeration.

Table 3a presents ordinary least-squares estimates of our model. As discussed above, we estimate separate models for each level of geography—zipcode, county, and state—and for each set of labor-market pooling proxies—net shipments per worker, managerial share of workers, and education. In total, therefore, the table presents nine regressions, three for each level of geography.

A set of results in Table 3a that warrants immediate discussion are the adjusted R^2 -values for each of the models. These range from near zero at the zipcode level of roughly 7% at the state level. On the surface, this suggests that state-level agglomeration is more closely related to agglomerative spillovers and natural advantages than are county- and zipcode-level agglomeration. This finding will prove robust in the analyses to follow. At the same time, the very low values for the adjusted R-squares suggest that our proxies for agglomerative spillovers and natural advantages explain only a fraction of the variation in agglomeration across industries. This raises the possibility that omitted industry attributes could bias our estimates.

To address that concern, Tables 3b and 3c provide a stringent set of robustness checks. Table 3b repeats the analyses in Table 3a but includes 2-digit SIC level fixed effects (20 in all), while Table 3c includes 3-digit SIC level fixed effects (140 in all).¹⁸ With these fixed effects added to the models, adjusted R^2 -values range from 4 to 21.5% with 2-digit fixed effects (Table 3b) and from 28 to 40% with 3-digit fixed effects (Tables 3c). Inclusion of these fixed effects, therefore, controls for a host of potentially important omitted determinants of agglomeration. But, at the same time, it is important to recognize that the fixed effects potentially soak up much of the meaningful variation in the data, making identification difficult, especially when 140 fixed effects are included in the model as in Table 3c. Bearing that tradeoff in mind, our discussion below emphasizes the OLS results in Table 3a but frequent references will also be made to the fixed-effects models as well.

An important result in Tables 3a, 3b, and 3c is the consistent evidence of a positive and significant influence of labor market pooling at all levels of

¹⁸ The latter model cannot be estimated when education is used to proxy labor-market pooling because the education variables are available up to the 3-digit level and, therefore, do not vary within the 4-digit subclassifications.

TABLE 3a
 OLS Gamma Regressions for Employment at All Establishments at the 4-Digit SIC Level^a

	Zipcode level			County level			State level		
	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers
Innovations from firms with fewer than 500 workers	-0.8664 -1.06	-1.1384 -1.36	-1.2112 -1.34	-1.8782 -1.72	-2.3535 -2.06	-2.4088 -1.96	-4.1396 -2.01	-4.3866 -2.08	-5.0579 -2.24
Innovations from firms with more than 500 workers	0.8874 1.17	0.9299 1.21	1.0587 1.32	0.7349 0.73	0.9231 0.88	1.0206 0.94	-0.7731 -0.41	-0.3629 -0.19	-0.3011 -0.15
Shipments net of inputs per worker	6.300×10^{-5} 3.27			1.553×10^{-4} 6.05			2.046×10^{-4} 4.24		
Managerial share of workers	0.0169 1.46			0.0210 1.33			-0.0123 -0.42		
Share of workers with Ph.D. or professional degree			-0.0096 -0.12			0.1149 1.04			0.0781 0.39
Share of workers with Master's degree			0.1285 2.47			0.1662 2.36			0.3776 2.92

Share of workers with Bachelor's degree	-0.0308	-0.0285	-0.1236
Manufactured inputs per \$ shipment	-1.23	-0.84	-1.98
Nonmanufactured inputs per \$ shipment	0.0069	0.0011	0.0953
Natural resources expenses per \$ shipment	0.49	0.06	2.69
Energy expenses per \$ shipment	-0.0319	-0.0740	-0.1652
Water expenses per \$ shipment	-1.39	-2.38	-2.89
Inventories per \$ shipment (nonperishability)	0.0050	0.0043	0.1174
R^2	0.33	0.21	3.10
Adj. R^2	-0.0449	-0.0816	0.0531
Sample size	-0.0336	-0.0838	-0.0486
	-0.62	-0.97	-0.31
	0.3308	0.6491	1.0779
	0.2614	0.5518	1.1939
	0.55	0.84	0.95
	0.0266	0.0230	0.1397
	1.27	0.79	2.53
	0.030	0.024	0.068
	0.005	0.004	0.050
	459	459	459
	427	427	427
	0.094	0.049	0.096
	0.076	0.024	0.072
	459	459	459

^a t -Ratios are below the coefficients. Constants are not reported in order to conserve space.

TABLE 3b
 SIC 2-Digit Fixed Effect Gamma Regressions for Employment at All Establishments at the 4-Digit SIC Level^a

	Zipcode level			County level			State level		
	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers
Innovations from firms with fewer than 500 workers	-0.4694 -0.56	-0.7951 -0.94	-0.5989 -0.65	-1.6353 -1.49	-2.0981 -1.82	-1.7462 -1.40	-2.8179 -1.42	-3.2661 -1.61	-3.0565 -1.42
Innovations from firms with more than 500 workers	1.1366 1.49	1.3517 1.75	1.3680 1.70	0.9726 0.97	1.3300 1.26	1.3867 1.27	0.1684 0.09	0.5195 0.28	0.5805 0.31
Shipments net of inputs per worker	1.049×10^{-4} 4.34			2.197×10^{-4} 6.89			2.196×10^{-4} 3.80		
Managerial share of workers	0.0421 2.96			0.0389 2.01			0.0357 1.05		
Share of workers with Ph.D. or professional degree			0.0477 0.50			0.1695 1.31			0.1647 0.73
Share of workers with Master's degree			0.2043 2.94			0.2003 2.14			0.5022 3.09

TABLE 3c
 SIC 3-Digit Fixed-Effect Gamma Regressions for Employment at All Establishments at the 4-Digit SIC Level^a

	Zipcode level		County level		State level	
	Net labor productivity	Managerial share of workers	Net labor productivity	Managerial share of workers	Net labor productivity	Managerial share of workers
Innovations from firms with fewer than 500 workers	-0.1646	-0.2501	-1.3302	-1.4147	-2.5586	-2.5291
Innovations from firms with more than 500 workers	-0.21	-0.32	-1.17	-1.22	-1.17	-1.14
Shipments net of inputs per worker	1.4545	1.5549	1.0547	1.1784	-0.1980	-0.1237
Managerial share of workers	2.16	2.26	1.06	1.16	-0.10	-0.06
	1.211×10^{-4}		1.784×10^{-4}		2.195×10^{-4}	
	3.67		3.66		2.34	
Manufactured inputs per \$ shipment	0.0062	-0.0116	-0.0049	-0.0271	0.0092	-0.0661
Nonmanufactured inputs per \$ shipment	-0.0180	-0.65	-0.0125	-1.04	0.17	-1.33
Natural resources expenses per \$ shipment	-0.72	-0.67	-0.17	-0.324	0.0736	-0.0245
Energy expenses per \$ shipment	0.0174	-0.0405	-0.0584	-0.0910	-0.1094	-0.1478
Water expenses per \$ shipment	0.84	-0.64	-1.58	-2.50	-1.54	-2.14
Inventories per \$ shipment (nonperishability)	0.0268	0.0011	0.0189	-0.0062	0.0736	0.0394
	0.35	0.06	0.62	-0.20	1.25	0.68
	-0.0076	-0.0206	0.0402	-0.0357	-0.0064	-0.1192
	-0.01	-0.27	0.36	-0.31	-0.03	-0.55
	0.0005	-0.0305	0.5661	0.5328	3.1269	3.0874
	0.02	-0.05	0.68	0.63	1.96	1.92
	0.594	-0.0155	0.0250	0.0034	0.1585	0.1387
	0.399	-0.65	0.71	0.10	2.35	2.07
Sample size	459	0.577	0.534	0.516	0.517	0.511
Number of fixed effects	140	0.374	0.312	0.285	0.286	0.277
		459	459	459	459	459
		140	140	140	140	140

^a *t*-Ratios are below the coefficients. Constants are not reported in order to conserve space.

geography, for all three proxies, and in both the OLS and fixed-effects specifications. The variable *Shipments net of inputs per worker* is always positive and significant in all of the models; *Managerial share of workers* is positive at the zipcode and county levels in the OLS and 2-digit fixed-effect models, though it is significant only for the 2-digit specification. Among the education variables, there is also a consistent pattern, with the *Master's degree* variable positive and at least marginally significant in all of the specifications. The consistency of these results provides strong evidence that labor market pooling is associated with industrial agglomeration. That finding is consistent with results from Dumais *et al.* (1997) who also report strong evidence of labor market pooling.

The coefficients on *Manufactured inputs* are positive but insignificant in the zipcode and county models, providing at most weak evidence that industries with a propensity toward input sharing concentrate at these levels of geography. The state-level coefficients, on the other hand, are all positive and significant in the OLS model (Table 3a), though significance is reduced continuously as one adds 2- and then 3-digit fixed effects to the model. Nevertheless, on balance, there is support for the idea that input sharing contributes to spatial agglomeration at the state level.

In contrast to the role of *Manufactured inputs*, the variable *Nonmanufactured inputs* has a negative coefficient in nearly all of the models and is significant at the state level for the OLS (Table 3a) and 3-digit fixed-effect specifications (Table 3c). Consistent with our priors, this suggests that the type of inputs upon which an industry depends influences the propensity to agglomerate. A reliance on manufactured inputs contributes to agglomeration. But, a reliance on service inputs—an important component of nonmanufactured inputs—does not, perhaps because these inputs are produced under constant returns or are not industry-specific and hence are available everywhere. Overall, our results on input sharing are in the spirit of Marshall [17].

There is also suggestive evidence for the importance of knowledge spillovers, but the evidence here is both mixed and weaker than for the other Marshallian microfoundations. At the county and state levels, *Innovations from firms with more than 500 workers* is nearly always insignificant and in some instances has a negative coefficient. However, at the zipcode level, large-firm innovation has a positive coefficient in all of the different models, with the coefficient not significant in the OLS specifications (Table 3a), marginally significant in the 2-digit fixed-effect specification (Table 3b), and significant in the 3-digit fixed-effect specification (Table 3c). On the other hand, small-firm innovation has consistently negative coefficients across the models, with the coefficient significant at higher levels of geography in the OLS specification. The result that large-firm innovation has a positive and significant effect only at the zipcode level is appealing given priors that knowledge spillovers attenuate

rapidly. But the negative coefficients on small firm innovations are difficult to explain, although these effects disappear with the inclusion of high-level fixed effects.¹⁹ On balance, therefore, we characterize our results here as suggesting that knowledge spillovers contribute to agglomeration at the local level, especially when innovative activity is based in large, well-established firms. But this conclusion should be viewed with caution, and further study is certainly warranted.

The remaining variables in Tables 3a, 3b, and 3c proxy for the importance of natural advantages as discussed earlier. On the input side, it is notable that industries that rely heavily on natural resources exhibit greater agglomeration only at the state level, with little effect at the zipcode and county levels. Specifically, the coefficients on the *Natural resources* variable are positive and significant at the state level but are insignificant at the other levels of geography. This result is quite apparent in the OLS and 2-digit fixed-effect models, but much less so in the 3-digit fixed-effect model. A similar result holds for reliance on *Water* related resources, which is also positive and significant in the 2- and 3-digit fixed-effect models, but not significant in the OLS model. In contrast, *Energy* is not significant in any of the models. Overall, these findings are consistent with priors, and they suggest that industries dependent on natural resources, such as timber and mining, are more likely to agglomerate because of a common need to locate close to the source of natural resource inputs. Moreover, as with reliance on manufactured inputs, reliance on natural resources contributes to agglomeration at the state level but is not evident at the zipcode and county levels.

The character of these findings is echoed in our estimates of the influence of product shipping costs on agglomeration. The variable *Inventories per \$ shipment* always has a positive and significant impact on state-level agglomeration, regardless of the choice of labor pooling proxy and regardless of the inclusion of industry fixed effects. This variable is always insignificant, however, at lower levels of geography. Given that *Inventories* is an inverse proxy for product shipping costs, these results support the idea that industries with output that is costly to transport are more likely to locate close to their markets and, as a result, exhibit less agglomeration.

¹⁹ Arguments from Saxenian [22], for example, suggest that knowledge generated at a given firm is more likely to spill over to the local economy if that knowledge is generated at small as opposed to large firms. In addition, our state-level results are somewhat at variance with Audretsch and Feldman [1], who found that industries with large expenditures on research and development were more likely to be concentrated at the state level. Of course, both the dependent and independent variables are different in our specification.

Taking all of these results together, an interesting pattern emerges. Reliance on manufactured and naturally occurring inputs and the production of perishable products serve to increase the importance of shipping costs in firm location decisions. That, in turn, positively affects state-level agglomeration but has little effect on agglomeration at lower levels of geography. In contrast, knowledge spillovers positively affect agglomeration at highly localized levels, while a reliance on skilled labor affects agglomeration at all levels of geography.²⁰

3.5. *The Geographic Nature of Agglomeration*

This section looks systematically at geographic differences in the determinants of agglomeration. We will focus on the degree to which the differences in the geography of agglomeration discussed above are statistically significant. In Tables 4a and 4b, we present OLS and 2-digit fixed-effect estimates of the difference in agglomeration at the county–zipcode level, $\gamma_c - \gamma_z$, and at the state–county level, $\gamma_s - \gamma_c$.²¹ Beginning once more with the adjusted R^2 -values, a different pattern from Table 3 emerges. First, the adjusted R^2 -values are very small in both tables for the county–zipcode regressions, ranging from 2 to roughly 9%. In addition, nearly all of the coefficients are individually insignificant in the county–zipcode regressions. This suggests that there is little systematic difference in the determinants of agglomeration at the county level relative to the zipcode level. In contrast, the adjusted R^2 -values are comparatively large for the state–county regressions, ranging from 8 to 9% for the OLS specification and from 27 to 29% in the 2-digit fixed-effects specification. These findings suggest that there is considerable systematic variation in the

²⁰ Two additional sets of robustness checks were carried out to evaluate the sensitivity of our findings to alternative specifications of the model. First, we experimented with using MSAs as the geographic unit of analysis. This was done in two ways: by estimating over MSAs only, discarding data from non-MSA locations, and treating each MSA as a separate geographic unit and by augmenting this sample with the non-MSA counties. Interpreted broadly, results from the MSA-only model are approximately a blend of those reported previously for the county- and state-level models. This is as anticipated since MSAs are larger than counties but smaller than states. Similarly, results from the MSA plus non-MSA county model are very similar to the county model. Again, this is as anticipated since the geographic scopes of the two models in this instance are similar. Details of these regressions are presented in Tables A-3a and A-3b in the Appendix.

A second set of robustness checks reestimated Tables 3a and 3b, measuring γ and the right-hand-side variables at the 3-digit SIC level. In general, results from those regressions support the principal findings presented above, with some variation. However, because the 4-digit models provide 459 industries while the 3-digit models aggregate to just 140 industries, the 4-digit models were favored. Results from the 3-digit-level analyses are not provided in order to conserve space.

²¹ Estimates from the 3-digit fixed-effect model are generally weaker but do not change the basic conclusions below and are not reported in order to conserve space.

TABLE 4a
 OLS Gamma Difference Regressions—All Establishments, 4-Digit SIC Level^a

	County–zipcode γ			State–county γ		
	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers
Innovations from firms with fewer than 500 workers	-1.0119 -1.59	-1.2151 -1.82	-1.1977 -1.69	-2.2613 -1.55	-2.0331 -1.38	-2.6491 -1.66
Innovations from firms with more than 500 workers	-0.1525 -0.26	-0.0067 -0.01	-0.0381 -0.06	-1.5080 -1.11	-1.2860 -0.95	-1.3217 -0.94
Shipments net of inputs per worker	9.230×10^{-5} 6.19			4.930×10^{-5} 1.44		
Managerial share of workers		0.0041 0.45			-0.0333 -1.64	
Share of workers with Ph.D. or professional degree			0.1245 1.95			-0.0368 -0.26
Share of workers with Master's degree			0.0377 0.93			0.2113 2.31
Share of workers with Bachelor's degree			0.0023 0.12			-0.0951 -2.15
Manufactured inputs per \$ shipment	0.0066 0.63	-0.0088 -0.81	-0.0059 -0.53	0.0969 4.05	0.0780 3.26	0.0942 3.76
Nonmanufactured inputs per \$ shipment	-0.0207 -1.37	-0.0291 -1.85	-0.0421 -2.35	-0.0671 -1.93	-0.0685 -1.98	-0.0912 -2.26
Natural resources expenses per \$ shipment	-0.0082 -0.74	-0.0033 -0.29	-0.0007 -0.06	0.1182 4.66	0.1147 4.49	0.1131 4.22
Energy expenses per \$ shipment	-0.0360 -0.76	-0.0503 -1.00	-0.0367 -0.65	0.0803 0.73	0.0352 0.32	0.1347 1.06
Water expenses per \$ shipment	0.1528 0.41	0.2904 0.76	0.3184 0.81	0.4545 0.54	0.5260 0.62	0.5448 0.61
Inventories per \$ shipment (nonperishability)	0.0200 1.23	0.0077 0.46	0.0054 0.31	0.1230 3.29	0.1238 3.32	0.1172 3.00
R^2	0.096	0.019	0.045	0.099	0.100	0.113
R^2 adj.	0.077	0.000	0.019	0.081	0.082	0.090
Sample size	459	459	427	459	459	427

^at-Ratios are below the coefficients. Constants are not reported in order to conserve space.

TABLE 4b
SIC 2-Digit Fixed Effect Difference Regressions—All Establishments, 4-Digit SIC Level^a

	County–Zipcode Gamma			State–County Gamma		
	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers
Innovations from firms with fewer than 500 workers	– 1.1659 – 1.77	– 1.3030 – 1.90	– 1.1473 – 1.56	– 1.1826 – 0.87	– 1.1680 – 0.86	– 1.3103 – 0.90
Innovations from firms with more than 500 workers	– 0.1639 – 0.27	– 0.0217 – 0.04	0.0186 0.03	– 0.8042 – 0.65	– 0.8105 – 0.65	– 0.8061 – 0.63
Shipments net of inputs per worker	1.148×10^{-4} 6.01			-1.530×10^{-7} 0.00		
Managerial share of workers		– 0.0031 – 0.27			– 0.0032 – 0.14	
Share of workers with Ph.D. or professional degree			0.1218 1.59			– 0.0048 – 0.03
Share of workers with Master's degree			– 0.0040 – 0.07			0.3019 2.74
Share of workers with Bachelor's degree			0.0307 1.30			– 0.0869 – 1.85
Manufactured inputs per \$ shipment	0.0085 0.69	– 0.0067 – 0.53	– 0.0048 – 0.36	0.0459 1.81	0.0456 1.82	0.0534 2.05
Nonmanufactured inputs per \$ shipment	– 0.0029 – 0.17	– 0.0141 – 0.80	– 0.0258 – 1.27	– 0.0276 – 0.79	– 0.0274 – 0.79	– 0.0263 – 0.65
Natural resources expenses per \$ shipment	0.0107 0.77	– 0.0029 – 0.20	– 0.0041 – 0.28	0.0843 2.95	0.0839 2.96	0.0694 2.36
Energy expenses per \$ shipment	– 0.0057 – 0.10	– 0.0098 – 0.17	– 0.0006 – 0.01	0.0206 0.18	0.0187 0.16	0.0911 0.66
Water expenses per \$ shipment	0.1893 0.45	0.1478 0.34	0.4270 0.94	1.5431 1.78	1.5372 1.77	1.7616 1.95
Inventories per \$ shipment (nonperishability)	0.0032 0.18	– 0.0069 – 0.37	– 0.00127 – 0.66	0.1138 3.10	0.1145 3.11	0.1189 3.11
R^2	0.143	0.071	0.090	0.315	0.315	0.344
R^2 adjusted	0.087	0.011	0.021	0.271	0.271	0.294
Sample size	459	459	427	459	459	427
Number of fixed effects	20	20	20	20	20	20

^a*t*-Ratios are below the coefficients. Constants are not reported in order to conserve space.

determinants of state-level agglomeration relative to the determinants of agglomeration at lower levels of geography.

Focusing on the state–county regressions, results support the most clear-cut findings from the previous section. *Manufactured inputs*, *Natural resources*, and *Inventories* all have positive and highly significant effects in the OLS model (Table 4a) and at least marginally significant effects following the inclusion of 2-digit SIC fixed effects (Table 4b). In addition, the *Water expenses* variable also has a positive and marginally significant effect once the fixed effects are added to the model. As noted above, these variables proxy for the importance of locating close to output markets and to factor inputs that tend to be concentrated in a relatively small number of states. In contrast, *Nonmanufactured inputs* has a negative and marginally significant effect in the OLS models and negative but not significant effects in the 2-digit fixed-effect model. Observe also that the various proxies for labor-market pooling are insignificant in all of the models with the exception of *Masters* and *Bachelors degrees*, which have opposite signs. *Masters* has a positive effect and *Bachelors* has a negative effect. As discussed above, labor pooling was found to positively influence agglomeration at all levels of geography. It is not surprising, therefore, that reliance on skilled labor does not help to systematically explain differences in agglomeration at the different levels of geography.²²

3.6. *The Agglomeration of New Establishments*

The patterns of agglomeration that we have studied thus far reflect decades of economic decisions. It is interesting to compare those patterns to agglomeration arising from more recent decisions. Accordingly, in this section we measure agglomeration at the 4-digit level using employment at just those establishments that were 5 years old or younger.²³ An important initial finding is that for every level of geography the average γ for employment at new establishments is very similar to the average γ for all employment. At the state, county, and zipcode levels, the averages for all employment are $\bar{\gamma}_s = 0.0485$, $\bar{\gamma}_c = 0.0193$, and $\bar{\gamma}_z = 0.0101$. For new-establishment employment, the aver-

²² Note also that *Innovations* from both small and large firms is insignificant in all of the models.

²³ To our knowledge, this is the first time that anyone has measured the agglomeration of employment at such newly established enterprises.

ages are $\bar{\gamma}_s = 0.00384$, $\bar{\gamma}_c = 0.0177$, and $\bar{\gamma}_z = 0.0104$.²⁴ Thus far, it appears that new-establishment agglomeration is similar to all-establishments agglomeration.

Tables 5a and 5b present OLS and 2-digit SIC fixed-effect estimates of the determinants of new-establishment agglomeration using the same specification as in Tables 3a and 3b.²⁵ As before, the adjusted R^2 -values are very low for the OLS specification (Table 5a). In contrast to previous findings, however, the adjusted R^2 remains low even after inclusion of 2-digit SIC fixed effects, with values ranging from 0 to 3%. The immediate conclusion, therefore, is that, compared to the agglomeration of all establishments, agglomeration of employment at newly created establishments is not as strongly related to the Marshallian microfoundations of agglomerative spillovers and to natural advantages. This conclusion is further supported by examination of the individual coefficients in Tables 5a and 5b. While the qualitative patterns are often similar to results from Tables 3a and 3b, the level of significance for new-establishment agglomeration is substantially reduced, especially for state-level agglomeration.²⁶

There are two ways in which one might account for these results. First, new establishments could differ systematically from older establishments. This would be the case in a dynamic setting in which new establishments that choose suboptimal locations are more likely to fail. In that case, surviving establishments would be more likely to be clustered in patterns that reflect the forces and benefits of agglomeration economies and proximity to natural advantages.

A second interpretation is that the more random pattern of locations among newly established enterprises reflects a fundamental change in the tendency to agglomerate. Today's business environment is in some ways quite different from that of 50 years ago. This has led some to question whether cities will play the same crucial role in the next millennium that they have in the one just

²⁴ In addition, the median difference between γ based on new versus all employment is very close to zero for each level of geography.

²⁵ Results from 3-digit fixed-effect specifications do not change the general conclusions discussed below and are not presented in order to conserve space.

²⁶ The principal exception to this generalization is the *Inventories* variable, which is positive and significant for all levels of geography and for all specifications of the model. This may indicate that newly established enterprises are especially sensitive to the cost of shipping their product to market when choosing their locations.

TABLE 5a
 OLS Gamma Regressions for Employment at Under Age 5 Establishments at the 4-Digit SIC Level^a

	Zipcode level			County level			State level		
	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers
Innovations from firms with fewer than 500 workers	0.6345 0.66	0.5769 0.59	0.7226 0.69	-0.1316 -0.11	-0.2931 -0.24	-0.0757 -0.06	-3.3807 -1.26	-3.6422 -1.35	-2.9624 -1.05
Innovations from firms with more than 500 workers	-0.8512 -0.95	-0.8242 -0.92	-0.9549 -1.03	-0.5000 -0.44	-0.5276 -0.46	-0.6298 -0.53	1.3178 0.53	1.3595 0.55	-0.4635 -0.19
Shipments net of inputs per worker	2.040×10^{-5} 0.90			1.650×10^{-5} 0.57			6.080×10^{-5} 0.97		
Managerial share of workers		2.254×10^{-3} 0.17			1.395×10^{-2} 0.81			1.621×10^{-2} 0.43	
Share of workers with Ph.D. or professional degree			0.0767 0.81			0.1199 0.98			0.0024 0.01

Share of workers with Master's degree	0.0226	0.0363	0.1365
Share of workers with Bachelor's degree	0.37	0.46	0.83
Manufactured inputs per \$ shipment	0.0222	0.0083	0.0465
Nonmanufactured inputs per \$ shipment	0.77	0.22	0.59
Natural resources expenses per \$ shipment	0.0170	0.0557	0.1451
Energy expenses per \$ shipment	0.0495	0.0508	0.1290
Water expenses per \$ shipment	0.60	2.51	2.94
Inventories per \$ shipment (nonperishability)	-0.0025	-0.0416	-0.0428
R^2	-0.11	-1.42	-0.67
Adj. R^2	0.0029	0.0078	0.0805
Sample size	0.17	0.35	1.70
	0.0042	0.0104	0.0859
	0.24	0.45	1.80
	-0.0113	-0.0376	-0.3379
	-0.16	-0.40	-1.66
	-0.3147	0.1107	1.1039
	-0.56	0.15	0.71
	0.0535	0.0907	0.1413
	2.17	2.86	2.06
	0.016	0.046	0.044
	0.000	0.027	0.025
	459	459	459
	427	427	427

^at-Ratios are below the coefficients. Constants are not reported in order to conserve space.

TABLE 5b
 SIC 2-Digit Fixed Effect Gamma Regressions for Employment at Under Age 5 Establishments at the 4-Digit SIC Level^a

	Zipcode level			County level			State level		
	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers
Innovations from firms with fewer than 500 workers	0.4922 0.49	0.5394 0.53	0.7910 0.72	-0.0565 -0.04	0.0310 0.02	0.2709 0.19	-3.5946 -1.29	-3.6133 -1.30	-2.8157 -0.96
Innovations from firms with more than 500 workers	-1.0628 -1.15	-1.0576 -1.15	-1.0518 -1.10	-0.7160 -0.61	-0.7660 -0.65	-0.8056 -0.65	1.4767 0.58	1.4893 0.58	-0.4592 -0.18
Shipments net of inputs per worker	3.360×10^{-5} 1.12			-1.690×10^{-5} -0.44			6.230×10^{-6} 0.08		
Managerial share of workers		-2.011×10^{-2} -1.18			-1.491×10^{-2} -0.69			2.435×10^{-3} 0.05	
Share of workers with Ph.D. or professional degree			0.1423 1.24			0.1496 1.02			0.0055 0.02
Share of workers with Master's degree			-0.0236 -0.29			-0.0491 -0.46			0.2122 0.96

Share of workers with Bachelor's degree	0.0184	0.0120	0.0000	0.0611	0.0611	-0.0281	0.1284	0.1280	0.0059
Manufactured inputs per \$ shipment	0.97	0.64	0.0184	2.53	2.55	0.0680	2.45	2.47	0.1447
Nonmanufactured inputs per \$ shipment	0.0154	0.0131	0.0056	-0.0115	-0.0093	-0.0227	0.0104	0.0097	0.0202
Natural resources expenses per \$ shipment	0.59	0.51	0.18	-0.35	-0.28	-0.58	0.15	0.14	0.25
Energy expenses per \$ shipment	0.0118	0.0067	0.0088	0.0211	0.0205	0.0209	0.1088	0.1086	0.1154
Water expenses per \$ shipment	0.54	0.31	0.39	0.76	0.74	0.71	1.81	1.81	1.89
Inventories per \$ shipment (nonperishability)	0.0444	0.0322	0.0185	-0.0181	-0.0263	-0.0486	-0.1915	-0.1903	-0.1878
	0.51	0.37	0.18	-0.16	-0.24	-0.37	-0.80	-0.79	-0.68
	0.0581	0.0120	0.0396	0.7320	0.7088	0.7262	1.0201	1.0230	1.7423
	0.09	0.02	0.06	0.89	0.86	0.83	0.57	0.58	0.96
	0.0499	0.0510	0.0501	0.0926	0.0970	0.0999	0.1172	0.1162	0.1136
	1.83	1.86	1.74	2.66	2.78	2.71	1.55	1.54	1.48
R^2	0.046	0.046	0.050	0.090	0.091	0.099	0.090	0.090	0.100
R^2 adj.	0.000	0.000	0.000	0.031	0.031	0.031	0.031	0.031	0.031
Sample size	459	459	427	459	459	427	459	459	427
Number of fixed effects	20	20	20	20	20	20	20	20	20

^at-Ratios are below the coefficients. Constants are not reported in order to conserve space.

ended.²⁷ Additional research is needed to discriminate between these two competing explanations for our result.

4. CONCLUSION

This paper has considered an important but understudied question in the empirical literature on agglomeration: What are the microfoundations of agglomeration economies? Using zipcode-, county-, and state-level employment data for the fourth quarter of 2000, we compute the measure of agglomeration developed by Ellison and Glaeser [5]. The agglomeration measure is then matched with various industry characteristics that proxy for the importance of knowledge spillovers, labor-market pooling, input sharing, natural advantages that affect input shipping costs, and product shipping costs. We find evidence of the importance of all of these determinants of agglomeration.

We also uncover an interesting geographic pattern that may well reflect the idiosyncratic characteristics of each of the determinants. Variables that proxy physical input and product shipping costs—including reliance on natural resources, manufactured inputs, and production of nonperishable output—all positively affect state-level agglomeration but have little effect on agglomeration at lower levels of geography. The geographic scope of these effects suggests that state-level transportation modes (i.e., train, truck, and barge transport) may play an important role in the location patterns of industries sensitive to shipping costs. At the other extreme, knowledge spillovers positively affect agglomeration only at the zipcode level, possibly because such spillovers attenuate rapidly across space. Finally, reliance on skilled labor positively affects agglomeration at all geographic levels. This latter result is particularly robust and may reflect spillover benefits that arise when skilled workers can seek out new job opportunities without having to move out of county or out of state. Together, these patterns explain an important share of the variation in state- versus county-level agglomeration across industries (up to 30%). Nevertheless, considerable unexplained variation in agglomeration remains, suggesting a role for continued research in this area.

We also find that employment at newly formed establishments is much less systematically related to the microfoundations of agglomeration than is employment at existing establishments. This could reflect a dynamic selection mecha-

²⁷ As Glaeser [8] notes, there are many factors that will come together to determine the future role of cities. One of these is the importance of agglomeration economies. If our findings can be interpreted to indicate that new firms agglomerate less and are less sensitive to Marshallian factors, then this would suggest a decline in the importance of cities. It is important to recognize, however, that there is a body of other evidence suggesting that agglomeration economies continue to exert powerful attractions, even to new establishments (see Rosenthal and Strange [20]).

nism, where only establishments that choose locations conducive to agglomerative spillovers and benefits from natural advantages survive. But our results could also reflect a fundamental change in the nature of establishment location decisions. Once again, further research is warranted.

APPENDIX

TABLE A-1
SIC Codes Used to Create Natural resources per \$ Shipment
from the 1992 BEA Input–Output Tables^a

Industry code	Description of industry category
10200	Poultry and eggs
10301	Meat animals
10302	Miscellaneous livestock
20100	Cotton
20201	Food grains
20202	Feed grains
20203	Grass seeds
20300	Tobacco
20401	Fruits
20402	Tree nuts
20501	Vegetables
20502	Sugar crops
20503	Miscellaneous crops
20600	Oil bearing crops
20701	Forest products
20702	Greenhouse and nursery products
30001	Forestry products
30002	Commercial fishing
40001	Agricultural, forestry, and fishery services
40002	Landscape and horticultural services
50001	Iron and ferroalloy ores, and miscellaneous metal ores, n.e.c.
60100	Copper ore
60200	Nonferrous metal ores, except copper
90001	Dimension, crushed and broken stone
90002	Sand and gravel
90003	Clay, ceramic, and refractory minerals
90004	Nonmetallic mineral services and miscellaneous
100000	Chemical and fertilizer minerals

^a Coal (Industry Code 07) and crude petroleum and natural gas (Industry Code 08) were included as part of the *Energy* variable rather than *Natural resources*. The latter, in contrast, is comprised of output from mining, agriculture, etc., as indicated by the list above.

TABLE A-2

Industry Codes Implicitly Used to Create *Nonmanufactured input per \$ shipment*
from the 1992 BEA Input-Output Tables^a

(32 Additional Categories Related to Government Services Are Omitted to Conserve Space).

Industry code	Description of industry category	Industry code	Description of industry category
650100	Railroads and related services	750003	Automobile parking and car washes
650200	Local and suburban transit and interurban highway passenger	760101	Motion picture services and theaters
650301	Trucking and courier services, except air	760102	Video tape rental
650302	Warehousing and storage	760201	Theatrical producers (except motion picture), bands, orchestras
650400	Water transportation	760202	Bowling centers
650500	Air transportation	760203	Professional sports clubs and promoters
650600	Pipelines, except natural gas	760204	Racing, including track operation
650701	Freight forwarders and other transportation services	760205	Physical fitness facilities and membership sports and
650702	Arrangement of passenger transportation	760206	Other amusement and recreation services
660100	Telephone, telegraph communications, and communication services	770100	Doctors and dentists
660200	Cable and other pay television services	770200	Hospitals
670000	Radio and TV broadcasting	770301	Nursing and personal care facilities
690100	Wholesale trade	770303	Other medical and health services
690200	Retail trade, except eating and drinking	770304	Veterinary services
700100	Banking	770305	Other medical and health services
700200	Credit agencies other than banks	770401	Elementary and secondary schools
700300	Security and commodity brokers	770402	Colleges, universities, and professional schools
700400	Insurance carriers	770403	Private libraries, vocational schools, and educational services
700500	Insurance agents, brokers, and services	770501	Business associations and professional membership
710100	Owner-occupied dwellings	770502	Labor organizations, civic, social, and fraternal associations
710201	Real estate agents, managers, operators, and lessors	770503	Religious organizations
710202	Royalties	770504	Other membership organizations
720101	Hotels	770600	Job training and related services
720102	Other lodging places	770700	Child day care services
720201	Laundry, cleaning, garment services, and shoe repair	770800	Residential care
720202	Funeral service and crematories	770900	Social services, n.e.c.
720203	Portrait photographic studios, and other miscellaneous personal	780100	U.S. Postal Service

TABLE A-2—Continued

Industry code	Description of industry category	Industry code	Description of industry category
720204	Electrical repair shops	780200	Federal electric utilities
720205	Watch, clock, jewelry, and furniture repair	780500	Other federal government enterprises
720300	Beauty and barber shops	790100	State and local government passenger transit
730101	Miscellaneous repair shops	790200	State and local government electric utilities
730102	Services to dwellings and other buildings	790300	Other State and local government enterprises
730103	Personnel supply services	800000	Noncomparable imports
730104	Computer and data processing services	810001	Scrap
730106	Detective and protective services	810002	Used and secondhand goods
730107	Miscellaneous equipment rental and leasing	820000	General government industry
730108	Photofinishing labs and commercial photography	830001	Rest of the world adjustment to final uses
730109	Other business services	840000	Household industry
730111	Management and public relations services	850000	Inventory valuation adjustment
730112	Research, development, and testing services, except noncommercial	880000	Compensation of employees
730200	Advertising	890000	Indirect business tax and nontax liability
730301	Legal services	900000	Other value added
730302	Engineering, architectural, and surveying services	910000	Personal consumption expenditures
730303	Accounting, auditing and bookkeeping, and miscellaneous services,	920000	Gross private fixed investment
740000	Eating and drinking places	930000	Change in business inventories
750001	Automotive rental and leasing, without drivers	940000	Exports of goods and services
750002	Automotive repair shops and services	950000	Imports of goods and services

^a *Nonmanufactured inputs* was calculated as a residual by forming $1 - \text{Value added} - \text{Manufactured inputs} - \text{Natural resources} - \text{Energy} - \text{Water}$, where all of these variables are measured per dollar of shipment. *Value added* (deflated by shipments) was obtained from the *Annual Survey of Manufactures* (ASM), while the other variables were obtained directly from the BEA input-output tables, the root source of which is the *Census of Manufactures* (CM). Because the ASM is based on a subset of the universe of the manufacturing establishments, while the CM covers the entire universe, some discrepancies occur. This and an imperfect matching of some of the BEA input-output industry codes to the 4-digit SIC code classifications account for a small number of industries for which our calculated *non-manufactured inputs* variable is negative.

TABLE A-3
 SIC 2-Digit Fixed Effect Gamma Regressions for MSA Measures of Geography—
 Employment at All Establishments at the 4-Digit SIC Level^a

	MSAs only			MSAs plus non-MSA counties		
	Net labor productivity	Managerial share of workers	Education of workers	Net labor productivity	Managerial share of workers	Education of workers
Innovations from firms with fewer than 500 workers	-2.5334 -1.43	-2.7657 -1.52	-2.4749 -1.25	-1.9882 -1.75	-2.4475 -2.05	-2.1143 -1.65
Innovations from firms with more than 500 workers	0.8681 0.54	1.1438 0.69	1.3789 0.80	1.0178 0.98	1.3751 1.26	1.4812 1.32
Shipments net of inputs per worker	2.421×10^{-4} 4.71			2.217×10^{-4} 6.71		
Managerial share of workers		-0.0194 -0.63			0.0376 1.87	
Share of workers with Ph.D. or professional degree			0.4812 2.34			0.1854 1.39
Share of workers with Master's degree			0.1291 0.87			0.2176 2.25
Share of workers with Bachelor's degree			-0.1137 -1.80			0.0158 0.38
Manufactured inputs per \$ shipment	0.0442 1.33	0.0105 0.31	0.0058 0.17	0.0126 0.59	-0.0114 -0.52	-0.0110 -0.48
Nonmanufactured inputs per \$ shipment	-0.0203 -0.45	-0.0433 -0.93	-0.0918 -1.68	-0.0026 -0.09	-0.0261 -0.85	-0.0503 -1.42
Natural resources expenses per \$ shipment	0.0885 2.37	0.0582 1.54	0.0338 0.85	0.0186 0.77	-0.0026 -0.11	-0.0157 -0.61
Energy expenses per \$ shipment	0.0235 0.15	0.0076 0.05	0.0364 0.20	-0.0079 -0.08	0.0093 0.09	0.0015 0.01
Water expenses per \$ shipment	1.5195 1.35	1.4083 1.22	1.8511 1.52	1.0142 1.40	1.0142 1.33	1.5559 1.96
Inventories per \$ shipment (nonperishability)	0.0158 0.33	-0.0029 -0.06	-0.0163 -0.32	0.0054 0.18	-0.0227 -0.70	-0.0257 -0.77
R^2	0.170	0.128	0.152	0.204	0.127	0.157
R^2 adj.	0.116	0.071	0.088	0.152	0.070	0.093
Sample size	459	459	427	459	459	427
Number of fixed effects	20	20	20	20	20	20

^at-Ratios are below the coefficients. Constants are not reported in order to conserve space.

REFERENCES

1. D. B. Audretsch and M. Feldman, R&D spillovers and the geography of innovation and production, *American Economic Review*, **86**, 630–640 (1996).
2. D. B. Audretsch, G. van Leeuwen, B. Menkveld, and R. Thurik, Market dynamics in the Netherlands: Competition policy and the response of small firms, *International Journal of Industrial Organization*, forthcoming.
3. A. Ciccone and R. E. Hall, Productivity and the density of economic activity, *American Economic Review*, **86**, 54–70 (1996).
4. G. Dumais, G. Ellison, and E. Glaeser, Geographic concentration as a dynamic process, NBER Working Paper 6270 (1997).
5. G. Ellison and E. Glaeser, Geographic concentration in U.S. manufacturing industries: A dashboard approach, *Journal of Political Economy*, **105**, 889–927 (1997).
6. G. Ellison and E. Glaeser, The geographic concentration of an industry. Does natural advantage explain agglomeration, *American Economic Association Papers and Proceedings* **89**, 311–316 (1999).
7. E. Glaeser, Learning in cities, *Journal of Urban Economics*, **46**, 254–277 (1999).
8. E. Glaeser, Are cities dying?, *Journal of Economic Perspectives*, **12**, 139–160 (1998).
9. E. Glaeser, H. Kallal, J. Scheinkman, and A. Shleifer, Growth in cities, *Journal of Political Economy*, **100**, 1126–1152 (1992).
10. G. Goldstein and T. Gronberg, Economies of scope and economies of agglomeration, *Journal of Urban Economics*, **16**, 91–104 (1984).
11. R. W. Helsley and W. C. Strange, Agglomeration economies and matching in a system of cities, *Regional Science and Urban Economics*, **20**, 189–212 (1990).
12. J. V. Henderson, Efficiency of resource usage and city size, *Journal of Urban Economics*, **19**, 47–70 (1986).
13. J. V. Henderson, A. Kuncoro, and M. Turner, Industrial development in cities, *Journal of Political Economy*, **103**, 1067–1085 (1995).
14. A. Jaffe, M. Trajtenberg, and R. Henderson, Geographic localization of knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics*, **108**, 577–598 (1993).
15. S. Kim, Regions, resources and economics geography: Sources of U.S. regional comparative advantage, 1880–1987, *Regional Science and Urban Economics*, **29**, 1–32 (1999).
16. P. Krugman, “Geography and Trade,” MIT Press, Cambridge, MA (1991).
17. A. Marshall, “Principles of Economics,” Macmillan, London (1920).
18. R. L. Moomaw, Productivity and city size: A critique of the evidence, *Quarterly Journal of Economics*, **96**, 675–688 (1981).
19. R. Nakamura, Agglomeration economies in urban manufacturing industries: A case of Japanese cities, *Journal of Urban Economics*, **17**, 108–124 (1985).
20. S. Rosenthal and W. C. Strange, Geography, industrial organization, and agglomeration, Center for Policy Research Working Paper 14, Syracuse University (1999).
21. J. Quigley, Urban diversity and economic growth, *Journal of Economic Perspectives*, **12**, 127–138 (1998).
22. A. Saxenian, “Regional Advantage: Culture and Competition in Silicon Valley and Route 128,” Harvard University, Cambridge, MA (1994).
23. L. Sveikauskas, The productivity of cities, *Quarterly Journal of Economics*, **89**, 393–413 (1975).