The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?

By GLENN ELLISON AND EDWARD L. GLAESER*

Scholars in many fields of economics have become very interested in Silicon Valley-style agglomerations of individual industries (J. Vernon Henderson, 1988; Michael E. Porter, 1990; Paul Krugman, 1991). These agglomerations are striking features of the economic landscape and may provide insights into the nature of the increasing-returns technologies and spillovers that are thought by many to be behind endogenous growth and business cycles.

In our previous work (Ellison and Glaeser, 1997), we noted that agglomerations may arise in two ways. In addition to explanations based on localized industry-specific spillovers, there is a simpler alternative: an industry will be agglomerated if firms locate in areas that have natural cost advantages. For example, the wine industry (the second most agglomerated industry in our study) is surely affected by the suitability of states' climates for growing grapes. If firms' location decisions are highly sensitive to cost differences (as found by Dennis Carlton [1983], Timothy J. Bartik [1985], and Henderson [1997], among others), then natural advantages may account for a substantial portion of observed geographic concentration.

In this paper we use the term "natural advantage" fairly broadly. Some possible examples would be that none of the more than 100,000 shipbuilding workers in the U.S. in 1987 worked in Colorado, Montana, or North Dakota and that the highest concentration of aluminum production (which uses electricity intensively) is in Washington (which has the lowest electricity prices). We will also speak of the concentration of the rubber and plastic footwear industry in North Carolina, Florida, and Maine and its absence from Alaska and Michigan as possibly reflecting natural advantages in the labor market. The industry is an intensive user of unskilled labor and faces tremendous competition from imports; hence we would expect to see it locate in low-wage states.

The simplest way to find effects of natural advantages on industry locations is to regress each industry's state-level employment on states' resource endowments as in Sukkoo Kim (1999). A problem with this approach, however, is that one can easily think of more potential advantages than there are states in the United States and fit each industry's employment distribution perfectly. We identify effects of natural advantages without overfitting by imposing cross-industry restrictions requiring the sensitivity of location decisions to the cost of a particular input to be related to the intensity with which the industry uses the input.

Using such an approach to estimate the effects of natural advantage on the 1987 locations of four-digit manufacturing industries, we find that industry locations are related to resource and labor-market natural advantages. Our primary goal is to see how much of the geographic concentration of industries reported in Ellison and Glaeser (1997) can be

† *Discussants:* Yannis Ioannides, Tufts University; Thomas Nechyba, Stanford University; Sukkoo Kim, Washington University–St. Louis.

* Ellison: Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, and NBER; Glaeser: Department of Economics, Harvard University, Cambridge MA 02138, and NBER. We thank the National Science Foundation for support through grants SBR-9515076 and SES-9601764. Ellison was also supported by a Sloan Research Fellowship. We thank Yannis Ioannides for valuable comments and David Hwang, Steve Jens, Alan Sorensen, and Marcus Stanley for research assistance.

attributed to natural advantages. In our preferred specification, we attribute about one-fifth of the concentration to observable natural advantages. Given that we are using only a small number of variables that capture advantages very imperfectly, we would guess that at least half of the concentration reported in Ellison and Glaeser (1997) is due to natural advantages. Nonetheless, there remain a number of highly geographically concentrated industries in which interfirm spillovers seem important.

I. Measuring Geographic Concentration

Following Ellison and Glaeser (1997), assume that industry i consists of K plants that sequentially choose locations from the set of states S in order to maximize profits. Suppose that the profits received by plant k when located in state s, π_{iks}, are given by

$$\log \pi_{iks} = \log \pi_{is} + g_s(v_1, \ldots, v_{k-1}) + \eta_{is} + \varepsilon_{iks}$$

where π_{is} is the expected profitability of locating in state s given observed state-specific costs, the function g_s reflects the effect on the profitability of state s of spillovers given that plants 1, ... , $k - 1$ have previously chosen locations $v_1, \ldots, v_{k-1}$, η_{is} is an unobserved common random component of the profitability of locating in state s, and ε_{iks} is an unobserved firm-specific shock.

In our model, plants in an industry may cluster relative to aggregate activity for three reasons: (i) more plants will locate in states with observed cost advantages; (ii) more plants will locate in states with unobserved cost advantages; and (iii) plants will cluster if spillovers are geographically localized.

Equilibrium geographic concentration is easy to compute for a particular specification in which the importance of unobserved natural advantages and spillovers are captured by parameters γ^{na} and $\gamma^{\mathrm{s}} \in [0, 1]$. The importance of unobserved natural advantages is reflected in the variance of η_{is}. Specifically, we assume that η_{is} is such that $2(1 - \gamma^{\mathrm{na}})\pi_{is}e^{\eta_{is}}/\gamma^{\mathrm{na}}$ has a χ^2 distribution with $E(\pi_{is}e^{\eta_{is}}) = \pi_{is}$ and $\mathrm{Var}(\pi_{is}e^{\eta_{is}}) = \gamma^{\mathrm{na}}\pi_{is}/(1 - \gamma^{\mathrm{na}})$. Spillovers are of an all-or-nothing variety: with probability γ^{s} a "crucial spillover" exists between each pair of plants. If such a spillover exists between plant k and plant ℓ, then plant k receives negative infinity profits if it does not locate in the same state as plant ℓ; otherwise its profits are independent of ℓ's location.

When the plants choose locations to maximize profits in this model, the expected share of employment located in state s is $E(S_{is}) \equiv \hat{S}_{is} = \pi_{is}/\Sigma_{s'}\pi_{is'}$. An index of geographic concentration beyond that accounted for by observed natural advantage is

$$\tilde{\gamma} \equiv \frac{\sum_s (S_{is} - \hat{S}_{is})^2/(1 - \sum_s \hat{S}_{is}^2) - H}{1 - H}$$

where H is the Herfindahl index of the plants' shares of industry employment. The property that makes this an appealing index (which follows from the same argument as in Ellison and Glaeser [1997]) is that $E(\tilde{\gamma}) = \gamma^{\mathrm{na}} + \gamma^{\mathrm{s}} - \gamma^{\mathrm{na}}\gamma^{\mathrm{s}}$.

The concentration index will reflect both unobserved natural advantages and localized spillovers. The index in Ellison and Glaeser (1997) is simply this index with the crude model where $\hat{S}_{is}$ is assumed to equal state s's share of overall manufacturing employment. As we better account for the effects of natural advantages on state-industry shares, we would expect the index of concentration to get smaller.

II. Does Natural Advantage Affect Industry Location?

In our empirical work we assume that average state-industry profits are

$$\log \pi_{is} = \alpha_0 \log(\mathrm{pop}_s) + \alpha_1 \log(\mathrm{mfg}_s) - \delta_i \sum_{\ell} \beta_\ell y_{\ell s} z_{\ell i}$$

where pop_s and mfg_s are the shares of total U.S. population and manufacturing employment in state s, ℓ indexes inputs to the production process, $y_{\ell s}$ is the cost of input ℓ in state s, and $z_{\ell i}$ is the intensity with which industry i uses input ℓ. The specification includes multiplicative industry dummies, δ_i, to account for the fact that observed cost differences will affect location decisions more in some industries than in others, both because of differences in the magnitude of the plant-specific shocks and because of transportation costs. For example, while one can easily imagine the fur industry concentrating in response to moderate cost differences, it seems hard to imagine that the concrete industry would concentrate geographically even if there were enormous differences in the costs of rocks and coal. In the estimation, the multiplicative industry effects are constrained to be nonnegative. Note that we have economized on the number of parameters by assuming that the effect on industry profitability of the difference in the cost of a particular input is proportional to the intensity with which the industry uses the input, rather than estimating a separate coefficient for each input for each industry.

Given this specification of the effects of natural advantages, expected state-industry employment shares are

$$E(S_{is}) = \frac{\text{pop}_s^{\alpha_0}\,\text{mfg}_s^{\alpha_1}\exp(-\delta_i \sum_\ell \beta_\ell y_{\ell s} z_{\ell i})}{\sum_{s'} \text{pop}_{s'}^{\alpha_0}\,\text{mfg}_{s'}^{\alpha_1}\exp(-\delta_i \sum_\ell \beta_\ell y_{\ell s'} z_{\ell i})}.$$

We estimate this relationship for the 1987 state employment shares of four-digit manufacturing industries by nonlinear least squares using 16 interactions designed to reflect advantages in natural resource, labor, and transportation costs. We normalize all of the interaction variables to have a standard deviation of 1 and choose their signs so that the β coefficients are expected to be positive. A larger estimated β value indicates that variation in the cost/intensity of use of that input has a larger effect in aggregate on the distribution of industries.

Results from estimating a base model without the multiplicative industry dummies are presented in Table 1A. The first column gives the name of the variable. Each is described in two lines: the first being the state-level input price variable (usually a proxy rather than a price) and the second being the industry-level variable reflecting intensity of use or the sensitivity of location decisions to input costs. The second column gives the coefficient, which should be interpreted as the percentage increase in the state's share of total industry employment caused by a one-standard-deviation increase in the explanatory variable; t statistics for the coefficient estimates are given in parentheses. Thus the coefficient of 0.170 for the variable electricity price × electricity usage means that the state's share of industry employment increases by about 17 percent (e.g., from 10 percent to 11.7 percent) with a one-standard-deviation increase in this variable. Table 1B shows the two industries for which the industry component of each variable is largest. For example, electricity is most important (as a fraction of value added) for primary aluminum and alkalies and chlorine. Table 1B also shows the states where the input cost is lowest (Washington, Idaho, and Montana for electricity) and the state where it is highest (Rhode Island).

The first six variables in the table (a–f) are designed to reflect the costs of six common inputs: electricity, natural gas, coal, agricultural products, livestock products, and lumber. All are highly significant and of the expected sign. The coefficients on several of these variables are among the largest we find, indicating that these variables reflect a substantial component of natural advantage.

The next six variables (g–l) relate to labor inputs. The first three are the average manufacturing wage in the state interacted with (g) wages as a share of value added, (h) the fraction of industry output that is exported, and (i) the fraction of U.S. consumption of the output good that is imported. Interactions (h) and (i) examine whether industries that are more competitive internationally are more wage-sensitive. All of these variables, however, will not matter if average wage differences are attributable to differences in labor productivity. Interactions (g) and (i) are significant and have the expected positive sign, but the coefficients are fairly small. We find no evidence of exporting industries concentrating in low-wage states.

The other labor input variables are designed to capture differences in the relative prices of different types of labor. Variable (j) is the interaction of the share of the adult population in the state without a high-school degree with

TABLE 1—EFFECT OF "NATURAL ADVANTAGES" ON STATE-INDUSTRY EMPLOYMENT

A. State variable × industry variable	Coefficient (t statistic)
(a) Electricity price × electricity use	0.170 (17.62)
(b) Natural gas price × natural gas use	0.117 (6.91)
(c) Coal price × coal use	0.119 (4.55)
(d) Percentage farmland × agricultural inputs	0.026 (2.58)
(e) Per capita cattle × livestock inputs	0.053 (5.08)
(f) Percentage timberland × lumber inputs	0.152 (11.98)
(g) Average mfg wage × wages/value added	0.059 (4.11)
(h) Average mfg wage × exports/output	−0.014 (−1.28)
(i) Average mfg wage × import competition	0.036 (3.10)
(j) Percentage without HS degree × percentage unskilled	0.157 (7.38)
(k) Unionization percentage × percentage precision products	0.100 (12.17)
(l) Percentage with B.A. or more × percentage executive/professional	0.170 (12.70)
(m) Coast dummy × heavy exports	−0.031 (−2.20)
(n) Coast dummy × heavy imports	0.017 (0.92)
(o) Population density × percentage to consumers	0.043 (3.68)
(p) (Income share − mfg share) × percentage to consumers	0.025 (4.49)

B. Variables[a]	Industries where most important (SIC)	Best states [worst state]
(a)	Primary aluminum (3334) Alkalies and chlorine (2812)	WA, ID, MT [RI]
(b)	Brick and clay tile (3251) Fertilizer (2873–4)	AK, LA, TX [HI]
(c)	Cement (3241) Lime (3274)	MT, NV, WY [VT]
(d)	Soybean oil (2075) Vegetable oil (2076)	NE, ND, SD [DC]

TABLE 1—*Continued.*

Variables[a]	Industries where most important (SIC)	Best states [worst state]
(e)	Milk (2026) Cheese (2022)	SD, NE, MT [MD]
(f)	Sawmills (2421) Wood preserving (2491)	AK, MT, ID [DC]
(g)	Industrial patterns (3543) Auto stampings (3465)	MS, NC, AR [MI]
(h)	Oil and gas machinery (3533) Rice milling (2044)	MS, NC, AR [MI]
(i)	Dolls (3942) Tableware (3263)	MS, NC, AR [MI]
(j)	Apparel (23) Textiles (22)	MS, KY, WV [AK]
(k)	Machine tools (354) Jewelry (391)	MI, NY, HI [SD]
(l)	Computers (357) (Periodicals) (2721)	DC, MA, CT [WV]
(m)	Rice milling (2044) Industrial gases (2813)	
(n)	Nonferrous metals (3339) Petroleum refining (2911)	
(o)	Potato chips (2036) Jewelry (3411)	DC, NJ, RI [AK]
(p)	Potato chips (2036) Jewelry (3411)	FL, CA, NY [NC]

[a]Letters in this column refer to state and industry variables in part A of the table.

the share of workers in the industry who are unskilled. Next, (k) is the interaction of unionization in the state (as a proxy for the presence of skilled workers) with the fraction of employees in the industry who are precision production workers. Variable (l) is the interaction of the fraction of the adult population in the state with bachelors' degrees or more education with the fraction of industry workers who are executives or professionals. All of these variables have a powerful positive effect.

The final four variables (m–p) relate to transportation costs. The first two (m and n) are designed to examine whether industries that are intensive importers or exporters of heavy goods tend to locate on the coast. Neither of the estimates is positive and significant.

The next two variables (o and p) are meant to capture the idea that firms will reduce transportation costs or improve their marketing by locating closer to their customers. They are interactions of the share of the industry's output that is sold to consumers with population density and with the difference between a state's share of income and its share of manufacturing employment. Both are significantly positively related to employment.

The coefficients on the natural advantages in specifications that include multiplicative dummies for two-digit and three-digit industries are similar. The tendency of labor-intensive industries to locate in low-wage states appears more pronounced in these regressions, while estimates of the effects due to unskilled labor, import competition, and income share minus manufacturing share become insignificant or negative.

III. Does Natural Advantage Explain Agglomeration?

Our greatest motivation for studying natural advantage is a desire to know whether it can account for a substantial portion of observed geographic concentration. Table 2 illustrates the effect on measured geographic concentration of accounting for observed natural advantages. Each row reports on the distribution of industry agglomeration indexes $\tilde{\gamma}$ obtained from a particular model of natural advantage. The first row describes the concentration index of Ellison and Glaeser (1997), which corresponds to the trivial model $E(S_{is}) = \text{mfg}_s$. The mean value of $\tilde{\gamma}$ in this model is 0.051. Only a few industries have negative $\tilde{\gamma}$'s. (This is noteworthy because the model has no transportation costs leading firms to spread out when serving local markets.) We regard the 28 percent of industries with $\tilde{\gamma} > 0.05$ as showing substantial agglomeration. For comparison, the $\tilde{\gamma}$ of the automobile industry (SIC 3711) is 0.127. Such extreme agglomeration is uncommon but far from unique: 12.8 percent of manufacturing industries have a $\tilde{\gamma}$ greater than 0.1.

The second row shows the results when we introduce the 16 cost/intensity of use interactions but do not allow industries to differ in the sensitivity of location decisions to observed cost differences. The mean $\tilde{\gamma}$ declines slightly to 0.048, and the overall distribution looks quite similar. The third and fourth rows describe the concentration indexes found when we allow for multiplicative dummies for each two- and three-digit industry, respectively. In these models, natural advantages have greater explanatory power, reducing the mean values of $\tilde{\gamma}$ to 0.045 and 0.041, respectively. We conclude that 20 percent of measured geographic concentration can be attributed to a few observable natural advantages.

TABLE 2—ESTIMATES OF RESIDUAL GEOGRAPHIC CONCENTRATION AFTER ACCOUNTING FOR OBSERVED NATURAL ADVANTAGES

Model	Mean $\tilde{\gamma}$	Percentage of industries with $\tilde{\gamma}$ in range <0.0	0.00–0.02	0.02–0.05	0.05–0.10	>0.1
A	0.051	2.8	39.9	29.2	15.3	12.8
B	0.048	3.9	39.9	30.1	13.7	12.4
C	0.045	3.1	42.9	29.4	13.5	11.1
D	0.041	4.4	42.9	29.8	13.3	9.6

Notes: Models A–D are different models of natural advantage: (A) no cost variables; (B) cost interactions introduced; (C) cost interactions plus dummies for two-digit industries; (D) cost interactions plus dummies for three-digit industries.

The fraction of industries that are extremely agglomerated in this measure declines, but only moderately: 9.6 percent of industries still have $\tilde{\gamma}$ greater than 0.1 in the latter specification. Another notable feature of the distribution of $\tilde{\gamma}$ is that the index is negative for only a very few industries. The finding that virtually all industries are at least slightly agglomerated is apparently fairly robust to the introduction of measures of cost advantages.

IV. Conclusion

Industries' locations are affected by a wide range of natural advantages. About 20 percent of observed geographic concentration can be explained by a small set of advantages. We think that this result is particularly notable given the limits on our explanatory variables. For example, nothing in our model can explain why there is no shipbuilding in Colorado, nor can it predict that soybean-oil production is concentrated in soybean-producing states, as opposed to being spread among all agricultural states. We hope that, in the future, others will provide better estimates than we have been able to give here. We conjecture that at least half of observed geographic concentration is due to natural advantages.

At the same time, there remain a large number of highly concentrated industries where it seems that agglomeration must be explained by localized intraindustry spillovers. Simple cost differences can not explain why the fur industry, the most agglomerated industry in our sample, is centered in New York. We see the attempt to provide a clearer understanding of the sources of these spillovers as an important topic for future research. Some results along these lines are described in Guy Dumais et al. (1997).

REFERENCES

Bartik, Timothy J. "Business Location Decisions in the United States: Estimates of the Effects of Unionization, Taxes, and Other Characteristics of States." *Journal of Business and Economic Statistics,* January 1985, *3*(1), pp. 14–22.

Carlton, Dennis. "The Location and Employment Decisions of New Firms: An Econometric Model with Discrete and Continuous Exogenous Variables." *Review of Economics and Statistics,* August 1983, *65*(3), pp. 440–49.

Dumais, Guy; Ellison, Glenn and Glaeser, Edward L. "Geographic Concentration as a Dynamic Process." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 6270, November 1997.

Ellison, Glenn and Glaeser, Edward L. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy,* October 1997, *105*(5), pp. 889–927.

Henderson, J. Vernon. *Urban development: Theory, fact, and illusion.* New York: Oxford University Press, 1988.

———. "The Impact of Air Quality Regulation on Industrial Location." *Annales d'Economie et de Statistique,* January–March 1997, (45), pp. 123–37.

Kim, Sukkoo. "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880–1987." *Regional Science and Urban Economics,* January 1999, *29*(1), pp. 1–32.

Krugman, Paul. *Geography and trade.* Cambridge, MA: MIT Press, 1991.

Porter, Michael E. *The competitive advantage of nations.* New York: Free Press, 1990.