5-2014

Simulating Confidence for the Ellison-Glaeser Index

Andrew J. Cassey
Washington State University

Ben O. Smith
University of Nebraska at Omaha, bosmith@unomaha.edu

Simulating Confidence for the Ellison-Glaeser Index*

Andrew J. Cassey and **Ben O. Smith**
School of Economic Sciences
Washington State University

January 2014

Abstract

The Ellison-Glaeser (1997) index is an unbiased statistic of industrial localization. Though the expected value of the index is known, *ad hoc* thresholds are used to interpret the extent of localization. We improve the interpretation of the index by simulating confidence intervals that a practitioner may use for a statistical test. In the data, we find cases whose index value is above the *ad hoc* threshold that are not statistically significant. We find many cases below the *ad hoc* threshold that are statistically significant. Our simulation program is freely available and is customizable for specific applications.

JEL classification: C63, L11, R14

Keywords: Ellison-Glaeser, localization, Herfindahl, simulation, confidence interval

Email: cassey@wsu.edu and tazz_ben@wsu.edu.
Correspondence: 101 Hulbert Hall, Pullman, WA 99164.
Department fax: 509 335 1173

*The authors thank Glen Ellison for providing data and two anonymous reviewers for outstanding suggestions. We also thank Ron Mittelhammer, Scott Colby, Julián Diáz, Tim Graciano, Tom Holmes, Robert Rosenman, Mykel Taylor, Mike Walrath, Ryan Bain, and seminar participants at the Bureau of Economic Analysis, Washington State University, and the 52nd Annual Meeting of the Western Regional Science Association.

1 Introduction

In many industries, employment is seemingly concentrated geographically beyond that of general economic or manufacturing activity, a phenomenon called localization. The theoretical literature has identified several reasons why this may occur. Localization could be due to natural (geographic or political) advantages such as extraction of oil in North Dakota, vineyards in Napa Valley, and casinos in Las Vegas. Localization also occurs without obvious natural advantage such as the auto industry in Michigan and the software industry in Silicon Valley. This may be due to spillovers from information, labor market pooling, or minimizing transportation costs in the supply chain.

Empirical tools have been developed to measure the extent of industrial localization by comparing industrial concentration to overall economic or manufacturing concentration. But some industries are composed of a small number of plants with large employment. Ellison and Glaeser (1997) first noted it is not desirable to consider such an industry localized only because of the small number of plants. They cite the U.S. vacuum cleaner industry (SIC 3635), where 75% of employment is in four large plants in different states, as an example of how we would not want to necessarily consider an industry localized just because 75% of industry employment is in four states.

Ellison and Glaeser develop an eponymous index, γ, that measures localization by controlling for overall manufacturing clustering and industrial concentration from small numbers, and whose values are comparable across industries and levels of geographic aggregation. Ellison and Glaeser show that if randomness is the only factor affecting localization—there are no natural advantages or spillovers—then the expected value of their index is zero. Therefore positive values of γ in the data indicate localization beyond that expected "had the plants in the industry chosen locations by throwing darts at a map" (p. 890). They then calculate γ for each of the 459 4-digit SIC manufacturing industries in the United States in 1987 and find the range of γ is between -0.013 and +0.630, with a median of 0.026 and a mean of 0.051. All but 13 industries have $\gamma > 0$. Though $\gamma > 0$ indicates industrial localization above that expected from pure randomness qualitatively, a more informative quantitative interpretation of γ is not obvious.

Consider the meat packaging industry (SIC 2011). Ellison and Glaeser calculate $\gamma = 0.042$.

This is obviously greater than zero, but is meat packaging very localized, somewhat localized, or barely localized? Ellison and Glaeser interpret their index by calculating the values for industries that are anecdotally thought to be agglomerated such as automobiles (SIC 3711), whose index value is 0.127, and carpet (SIC 2273), whose index value is 0.378. They also calculate γ for industries that seem anecdotally not to be localized such as miscellaneous concrete products (SIC 3272), whose index value is 0.012, and bottled and canned soft drinks (SIC 2086), whose index value is 0.005. Therefore, Ellison and Glaeser call industries with $\gamma > 0.050$ very localized, industries with $0.020 < \gamma \leq 0.050$ somewhat localized, and industries with $\gamma < 0.050$ barely localized. These *ad hoc* thresholds categorize 43% of industries as barely localized and 28% of industries as very localized.

After describing Ellison and Glaeser's (1997) model and index in section 2, in section 3 we improve the quantitative interpretation of the Ellison-Glaeser index by simulating confidence intervals. We write computer code that simulates the Ellison and Glaeser model in order to calculate how likely it is for an industry to achieve a value of $\gamma = c$ for any c as a matter of pure randomness. Our simulated confidence intervals depend on the number of plants in the industry and the standard deviation of the underlying lognormal plant employment distribution. Because the plant employment standard deviation is difficult to obtain or estimate from the data, we also provide confidence intervals based on the number of plants and the industry's plant Herfindahl. Using our confidence intervals, a practitioner can conduct a formal statistical test for localization using the Ellison-Glaeser index as the measure.

Section 4 reports the results from our simulation showing that confidence intervals increase in the standard deviation of the underlying logarithm of the plant employment distribution and asymptotically decrease to zero width in the number of plants in the industry. Therefore, a critical value is not a constant across all industries but rather varies depending on industry parameters. The same γ could indicate a statistically significant level of localization for one industry but not another. The reason is that though the expected value of γ does not depend on the number of darts thrown and the size of the darts (plant employment), the distribution of γ does.

In section 5 we test which manufacturing industries have a statistically significant level of localization. Our tests are performed on the same data used by Ellison and Glaeser (1997). We find that 78% of industries have a statically significant level of localization. We find 2 of 127 of

Ellison and Glaeser's very localized industries and 12 of 131 of their somewhat localized industries have levels of localization that are not statistically different from randomness at the 5% level. In addition, we find that 112 of the 201 industries they consider barely localized have a less than 5% chance of obtaining their level of localization randomly. We also apply our confidence intervals on the 6-digit NAICS data presented in Holmes and Stevens (2004) and find that there exists industries that are statistically diffuse.

That we find the *ad hoc* thresholds set by Ellison and Gleaser can lead to type I errors but frequently lead to type II errors (at the national level) is a matter of the thresholds set, but more importantly that as the number of plants becomes large, the chance of that industry achieving even a small positive γ becomes vanishingly small. Thus establishing any threshold by collecting a percentage of industries with a γ below that level will be subject to type II errors on those industries that have many more plants than other industries below the threshold. The same is true for industries whose plant employment distribution variance is smaller.

The computer code we use in our simulations is publicly available. It is written to be customizable so that a researcher can get the exact confidence interval for their application. (Our code also has the option to simulate confidence intervals for the similar measure of localization proposed by Maurel and Sédillot (1999) and can calculate confidence intervals for geographic weight modifications as in Ellison and Glaeser (1999).) The confidence interval tables we include here are just an illustration of the program output.

That until now there has been no quantitative interpretation of the Ellison-Glaser index is an important problem because it is a frequently used measure of industrial localization. To cite just a few examples, Rosenthal and Strange (2001) determine the underlying factors in agglomeration by regressing the Ellison-Glaeser index. Overman and Puga (2010) use the Ellison-Glaeser index to quantify gains from labor pooling while Gautier and Teulings (2003) focus on labor market density. Briant, Combes, and Lafourcade (2010) examine how different zoning systems can impact economic estimations, and Combes (2000) uses the index as justification for his modeling assumptions. About 2000 articles have cited Ellison and Glaeser (1997) and its working paper version (1994) according to Google Scholar as of May 2013. One reason for its popularity is that the data requirement to use the index is relatively low.

This paper is similar in spirit to Duranton and Overman (2005) who also simulate a confidence interval around a localization statistic in order to give statistical significance to empirical results. But Duranton and Overman do not base their localization statistic on the Ellison and Glaeser index. Rather they create their own index using the physical distance between plants. Though the Duranton and Overman statistic is more accurate, it is also far more difficult to obtain the data requirements of physical distance between plants. A more recent localization measure proposed by Billings and Johnson (2013) has a similar relatively high data requirement. Therefore we believe that our confidence interval for the Ellison and Glaeser index is useful for many research applications.

2 The Ellison-Glaeser Index

Ellison and Glaeser (1997) propose a model in which N plants in an industry sequentially choose to locate in one of M contiguous non-overlapping discrete regions. These regions are bins without internal distance and there is no notion of contiguity. Plants know their employment size, which is drawn from a lognormal distribution $X \sim \log \mathcal{N}(\mu, \sigma^2)$. (For convenience, we loosely refer to μ as the mean and σ as the standard deviation of X.) Let v_k denote the location of plant k. In the model, plant k chooses region i to maximize profit π_{ki}:

$$\log \pi_{ki} = \log \bar{\pi}_i + g_i(v_1, ..., v_{k-1}) + \varepsilon_{ki}$$

where $\bar{\pi}_i$ is the average profit in region i, g_i is the spillover indicating the profit obtained from plants 1 to $k-1$ also locating in i, and ε_{ki} is a plant's individual random component.

Let s_i be the share of industry employment in region i and x_i be the share of total manufacturing employment in region i. If spillovers and natural advantages are turned off in the model, then $g(\cdot) = 0$ and plants locate in the region with the highest average profit. Thus the likelihood of plant k locating in region i is x_i. Therefore, a measure of raw geographic concentration is the Gini statistic, $G = \sum_{i=1}^{M}(s_i - x_i)^2$.

Ellison and Glaeser (1997) show that when there is a small number of plants in the industry, clustering, as measured by G, can result from chance. Therefore they construct the following index:

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

where $H = \sum_{k=1}^{N} z_k^2$ is the plant Herfindahl index for that industry and z_k is plant k's share of industry employment. At high levels of aggregation, such as industrial sectors or all manufacturing, the number of plants is large and the Herfindahl index nears zero. Thus $\gamma = \frac{G}{1-\sum_i x_i^2}$ so that the Ellison-Glaeser index is simply a rescaled Gini statistic. But when the the number of plants is small, γ can greatly differ from G. Ellison and Glaeser show that $\mathbb{E}[\gamma] = 0$ when there are no natural advantages or spillovers. Positive values measure localization beyond that expected by pure randomness whereas negative values measure plants choosing to locate more diffusely than expected by randomness.

Ellison and Glaeser show the expected value of their statistic is robust to the level of geographic aggregation provided the pieces sum to the whole and the spillover function applies completely within a region and does not apply at all to any contiguous region. They write, "...the index is designed to facilitate comparisons across industries, across countries, or over time. When plants' location decisions are made as in the model, differences in the size of the industry, the size distribution of plants, or the fineness of the geographic data that are available should not affect the index" (p. 890).

The robustness of the expected value to the level of geographic aggregation is true in theory if spillovers are assumed to have a value of one within an arbitrary geographic region and zero otherwise. Feser (2000) shows that in practice, the Ellison-Glaeser index is not robust to geographic division because their spillover assumption is not realistic. A more realistic assumption is that spillover strength decays over physical distance without appealing to arbitrary region borders, as in Duranton and Overman (2005), although Kerr and Kominers (2010) argue that the spillover goes to zero after some distance. However, the data requirements for calculating the Duranton and Overman localization measure are relatively high, thus limiting its applicability in practice. We therefore believe it is of great practical and generalizable use to simulate the confidence interval for the Ellison-Glaeser index.

3 Simulation Set Up

To simulate a confidence interval for γ, we follow the set up in Ellison and Glaeser (1997) by using an employment weighted map of the U.S. states as the specification of the x_i from (1) and assuming the plant employment distribution of each industry is lognormal. We follow Ellison and Glaeser in using the lognormal distribution for plant employment because of empirical evidence such as that provided in Stanley, Buldyrev, Havlin, Mantegna, Salinger, and Stanley (1995) and Cabral and Mata (2003). The lognormal distribution requires two parameters to be specified: the mean and standard deviation from the corresponding normal distribution. For each simulation, we specify particular parameter values as well as the number of plants in the industry. Given the number of plants and underlying distribution, a pseudorandom number generator picks employment for each of the N plants in the industry from the lognormal distribution. A pseudorandom number generator also picks the location of each plant randomly from the distribution of non-farm employment in the data of the $\mathbf{x}$ vector. For this application, a run-of-the-mill pseudorandom number generator is biased. See appendix A for details of the pseudorandom number generator we use and why we use it.

Thus we give the model data on x_i and then calculate the share of industry employment in each region s_i and the plant Herfindahl for the industry H from the random draws of plant employment size and location. These are the three ingredients to calculate γ. We do this 100,000 times and then order the realizations of γ to create the empirical distribution function. We calculate the critical values for the intervals containing, for example, the middle 95% of the observations, as well as the p-values. We then change either the number of plants or one of the parameters of the lognormal distribution and repeat the process, thus creating confidence intervals as a function of three parameters. Because there are no natural advantages or spillovers in our simulation, each realization of γ is purely due to randomness. Thus, the expected value of γ is zero regardless of the parameters chosen for each simulation.[1] Our simulated confidence interval can then be used to test if a γ in the data could have been generated from randomness to some desired statistical level.

[1]Our program outputs the mean of the raw γ values, as well as other checks, in order to verify our simulation is correct.

Our program is freely available at: `http://goo.gl/n1N06`. It is customizable so that a practitioner can decide on a statistical level and simulate the confidence interval for a particular application. A user can change the geographic scope by inputting a different **x** vector than the non-farm employment of 50 U.S. states we use. A user can also incorporate different weights in the **x** vector to account for observed natural advantages as in Ellison and Glaeser (1999). In that case, the Ellison-Gleaser index is rescaled so that the expected value, given the inputted natural advantage, is zero and our simulated confidence intervals apply to that rescaling. Finally, the program has an option to generate the confidence intervals for the similar Maurel and Sédillot (1999) index of localization. For more information about how to install the program, see appendix B.

4 Results

Below we list a theorem and two generalized results obtained from our numerical simulations. Table C.1 in appendix C gives a brief sample of the critical values from the simulation. These critical values are calculated from the simulation specified in section 3 and as such do not consider mistakes in data entry or if the geographic space is continuous and has spillovers extending into other regions.

Theorem. *The confidence interval of the Ellision-Glaeser index does not depend on the mean, μ, of the logarithm of the plant employment distribution.*

Proof. The Ellison-Glaeser index is a function of x_i, s_i, and H. The lognormal distribution is used to randomly determine the plant size but not location. Therefore the x_i are taken as exogenous in (1) and do not depend on μ. We show that the plant employment share used in s_i and H do not depend on μ either, and thus the confidence interval for γ cannot depend on μ.

Let z_k be plant k's share of industry employment. Then $s_i = \sum_{k=1}^{N} z_k 1_i(k)$ and $H = \sum_{k=1}^{N} z_k^2$, where N is the number of plants overall in that industry and $1_i(k)$ is an indicator function specifying that plant k is in region i. Using the the inverse CDF of a lognormal distribution, a randomly generated plant size can be specified: $s_k = e^{\mu} e^{-\sqrt{2}\sigma Erfc^{-1}[2d_k]}$ where $Erfc^{-1}$ is the inverse complementary error function and d_k is a random draw from $(0, 1)$. The plant employment share is then:

$$z_k = \frac{s_k}{\sum_{j=1}^{N} s_j} = \frac{e^{\mu} e^{-\sqrt{2}\sigma Erfc^{-1}[2d_k]}}{\sum_{j=1}^{N} e^{\mu} e^{-\sqrt{2}\sigma Erfc^{-1}[2d_j]}} = \frac{e^{-\sqrt{2}\sigma Erfc^{-1}[2d_k]}}{\sum_{j=1}^{N} e^{-\sqrt{2}\sigma Erfc^{-1}[2d_j]}},$$

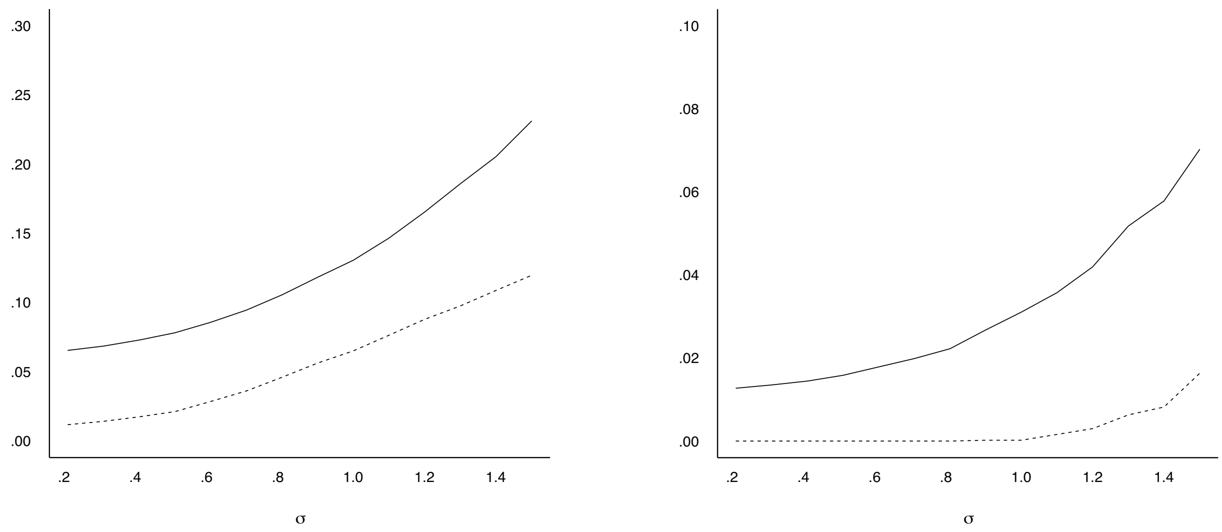
which does not depend on μ. □

In the lognormal distribution, μ functions as a scaling parameter. Given σ, changing μ simply rescales the distribution and thus does not result in any change to the index. Deltas (2003) has the same result in showing small sample bias in Gini coefficient estimates. The proof also makes clear that s_i and H do depend on σ. We turn to numerical simulations to see how σ affects the confidence interval of γ.

Result 1. *Increasing the standard deviation, σ, of the logarithm of the plant employment distribution increases the width of the confidence interval.*

The solid line in figure 1 shows the width of the confidence interval capturing the middle 95% of observations for a realistic domain of σ (Deltas 2003) while holding the number of plants in the industry fixed. Also graphed is the percent of observations that randomly have $\gamma > 0.05$, the value Ellison and Glaeser (1997) consider to be very localized. This dashed line may be thought of as the chance of a type I error. The left panel of figure 1 shows the results for 20 plants whereas the right panel shows the results for 100 plants. Table C.1 contains other values. While 20 plants is small compared to 300 plants, the median number in an industry nationally, 20 plants may not be small for applications on city or county data. Therefore, at the threshold of $\gamma = 0.05$, the chance of a type I error becomes quite big for large, but plausible, values of σ at the local level. The right panel showing 100 plants is more realistic on a national scale. Again the width of the confidence interval increases with σ. But with as many as 100 plants, there is little chance of a type I error for realistic values of σ.

These graphs are upward sloping indicating the width of the confidence interval and the chance of a Type I error increases with σ. The reason is because increasing the standard deviation increases the likelihood that there are large plants. Then these large plants are randomly assigned to a region. Therefore, statistically, it is more difficult to distinguish whether a large employment share is due to spillovers or a “fat” dart landing randomly.



(a) N=20

(b) N=100

Figure 1. Given a fixed number of plants ($N = 20$ in the left panel and $N = 100$ in the right panel), the width of the confidence interval increases with the standard deviation of the logarithm of the plant employment distribution. The solid line indicates the width of the confidence interval to capture 95% mass whereas the dashed line is the probability that $\gamma > 0.05$ will randomly occur when there are no natural advantages or spillovers. Note the change in vertical axis scale between the panels.

Result 2. *Increasing the number of plants N decreases the width of the confidence interval.*

Figure 2 shows how the confidence interval capturing the middle 95% of observations is downward sloping in the number of plants in the industry. In the figure, we set $\sigma = 0.6$, which is a realistic value for industries on a national scale (Deltas 2003). As can be seen, the width of the confidence intervals asymptotically approaches zero. For empirical purposes, the confidence interval width is almost zero when there are more than 500 plants in the industry, regardless of the (realistic) underlying employment distribution. Therefore the level of localization of industries that have a small but positive γ may be statistically significant at the 5% level if there are many plants.

As before, the dashed line graphs the percent of observations for which $\gamma > 0.05$ by chance. While there is about a 10% chance of a type I error at the Ellison and Glaeser threshold when there are only ten plants, there is essentially no chance of a type I error when the number of plants is greater than 100 for an industry with a plausible standard deviation in its plant employment distribution. The reason these graphs are decreasing is because clustering of a few plants could be due just to small numbers, whereas it is increasingly unlikely that many plants randomly locate in the same region.

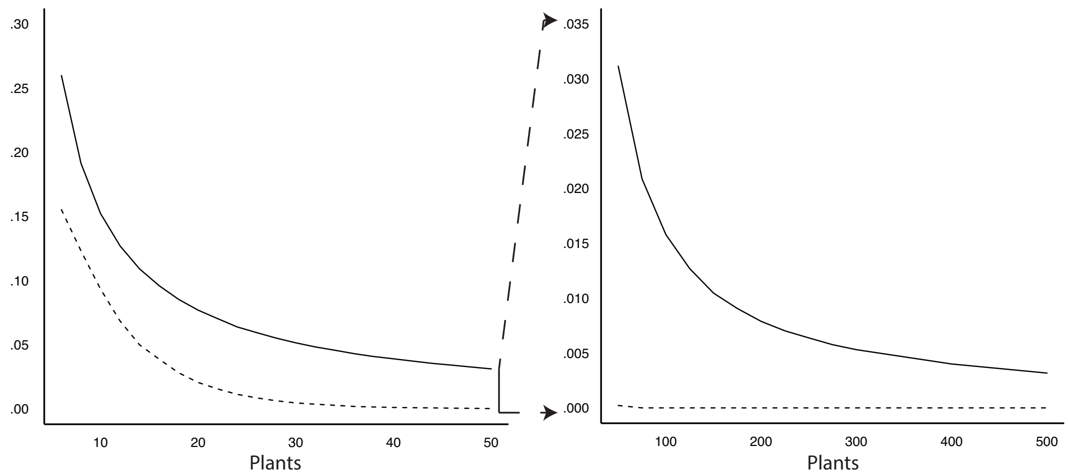


Figure 2. Given a fixed plant size distribution ($\sigma = 0.6$), the width of the confidence interval decreases with the number of plants in the industry. The solid line indicates the width of the confidence interval to capture 95% mass whereas the dashed line is the probability that $\gamma > 0.05$ will randomly occur when there are no natural advantages or spillovers. Note the change in scale (on both axes) between the panels.

Because we ran our simulation 100,000 times for each (N, σ), our critical values are very stable in the sense that if we ran another 100,000 runs on the same parameter values, the critical values would be very nearly identical to five decimal places. Even at a very low plant count such as ten and a relatively large σ such as one, our 95% critical value is statistically different from the the 0.02 threshold used by Ellison and Glaeser if it is outside of $[0.0197, 0.0203]$. For example, with $N = 10$ and $\sigma = 1$, our critical value of 0.095 is outside of that range, suggesting that in principle there is an important reason for a practitioner to do the extra work of simulating a confidence interval for a particular application rather than using a constant threshold. Whether this is important in practice depends on how often a researcher calculates γ for an industry with ten plants and $\sigma = 1$. Though this few of plants is not common in national applications, it is more common for local applications. Furthermore, for a nationally representative industry with 300 plants and $\sigma = 0.6$, our 95% critical value is 0.002, which is significantly less than the 0.020 threshold, indicating there is a very good chance of a type II error.

To get an idea of the chance of type I and type II errors using a constant threshold, see figure 3. That figure compares our 95% critical value (curved surface) to Ellison and Glaeser's (1997) 0.02 threshold (the flat plane). We see that at a low plant count and high σ, the *ad hoc* threshold leads to type I errors whereas a high plant count and low σ lead to type II errors. We numerically

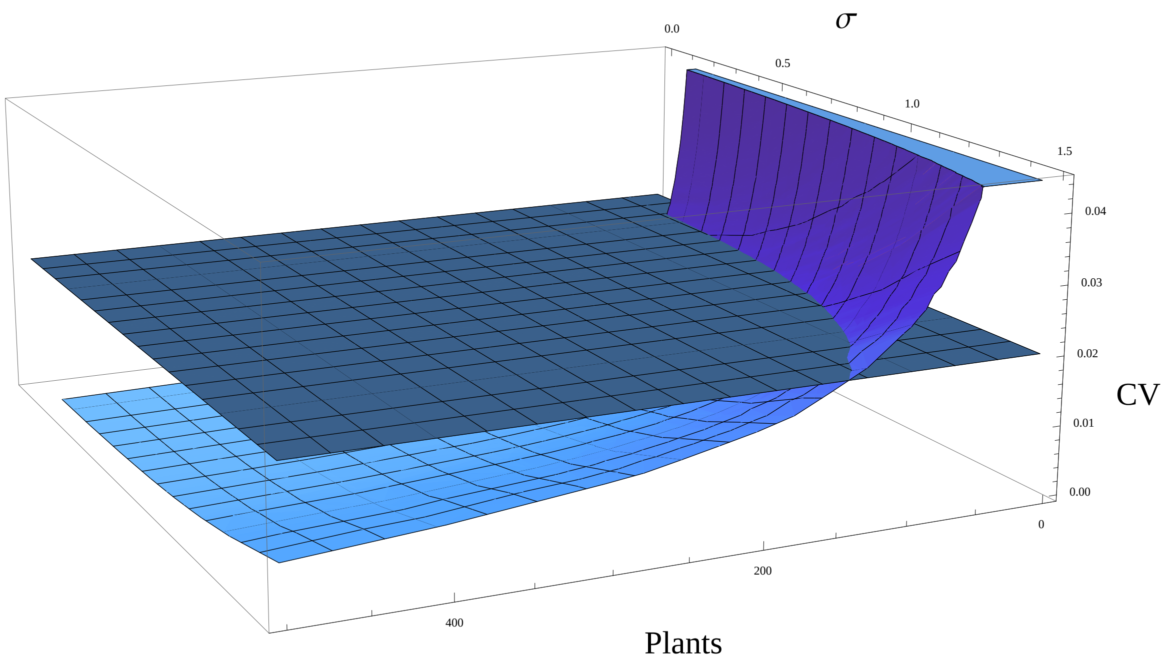


Figure 3. Comparison of the simulated 95% critical value to the 0.02 threshold for localization.

integrate the difference in these surfaces to get a quantitative measure of the importance of using a critical value that depends on industry parameters. If we assume there is a uniform distribution of γ on $[0.0, 0.1]$, then there is a 2.6% chance of a type I error and a 10.6% chance of a type II error when there are fewer than 500 plants.

4.1 Calculated Herfindahls

In principle, the confidence interval of γ depends on three parameters: μ, σ, and N. However, the confidence interval does not depend on μ. Therefore our critical values depend on σ and N. Since it is difficult for a researcher to obtain or estimate σ from the data, our work up to now has limited applicability. Therefore, as a matter for practice, we calculate critical values for γ as a function of the Herfindahl. We then map the practical parameters (N, H) to the actual parameters (N, σ).

The Herfindahl, however, is not unique in that the same H value is obtainable from different underlying σ values. Those different σs imply different critical values for γ. For this reason, we calculate Herfindahl critical regions based on our 100,000 runs. These regions indicate the range of H obtainable from a specific σ in 95% of observations. These ranges can overlap for different σs. Because the Herfindahl critical region is simply a functional transform of the underlying σ and N, it is not independent, and therefore does not add additional uncertainty to the critical values of γ.

In the appendix, table C.1 gives example output from our program showing the mapping between σ and the Herfindahl that a researcher may use to test the significance of a γ value they are analyzing. To use this table, the researcher would know the industry's plant Herfindahl and the number of plants, but not underlying the standard deviation of the employment distribution. They would first go to the row with their number of plants from the data. The researcher would scan over the 95% Herfindahl ranges generated by our program within that number of plants and settle on the Herfindahl ranges that match their data. A Herfindahl range implies the unknown lognormal employment distribution parameter σ. The corresponding γ critical values are the lower and upper bound for which 95% of our simulated random observations lie between. Thus in order for γ to be statistically significant, the value must be outside of this range. Therefore the researcher has options on how conservative to be in assigning statistical significance to the γ they are analyzing. The most conservative critical values would be the widest range of γ critical values whereas liberal critical values would be the narrowest range.

To see how a researcher could use our results, consider the following: A researcher is testing if lawn and garden equipment (SIC 3524) is localized nationally. There are 165 plants in this industry, the Herfindahl is 0.043, and $\gamma = 0.014$. If the researcher uses our program, they can input $N = 165$ into our program and specify if they want to use the entire range of simulated Herfindahl values or condition on a subrange. If they use table C.1, then they would first find the row for $Plants = 150$, which is nearest value in the table less than 165.[2] Of those rows, the researcher finds 0.043 is within the 95% Herfindahl range for two rows. The researcher then looks over to the 95% γ critical values and finds the the narrowest distribution of critical values is $[-.010, .013]$ while the largest range is $[-.014, .019]$. This largest range corresponds to the most conservative critical values for γ to be statistically significant at the 5% level. With $\gamma = 0.014$, Ellison and Glaeser (1997) classify this industry as not very localized. But since $\gamma = 0.014$ is greater than 0.013, there is at least

[2] Our table provided in this paper is a sample of the entire table found at `http://goo.gl/Ox7YD`.

one value of σ in which this industry could be considered to have a statistically significant level of localization. Since $0.014 < 0.019$, it is not the case that this industry has a statistically significant level of localization for any reasonable value of σ.

The practical usefulness of our simulation somewhat depends on whether the range of Herfindahls maps onto a narrow difference between the conservative and liberal critical values. The liberal confidence interval must be within the conservative confidence interval. When a calculated γ is within the narrow liberal confidence interval, then we know that there is no plausible value of σ which would cause the industry to have a statistically significant level of localization. Likewise when a calculated γ is outside the wide conservative confidence interval, there is no plausible σ that could achieve that level of localization from randomness. Thus the question is "How often are the calculated γs in between?"

Result 3. *The width in the range between the liberal and conservative critical values decreases with the number of plants.*

By 100 plants, the difference between the conservative and liberal confidence intervals is zero to two decimal places and by 400 plants the difference is zero to three decimals. Thus for industries with large plant counts, there is essentially no difference in these ranges and so the confidence intervals are particularly useful.

An alternative approach is to condition the simulation on a range of inputted Herfindahl values and back out from the simulation the largest σ that could generate any value in that Herfindahl range. For any number of plants in the industry, we assign the Herfindahl value from the data into a bin of similar Herfindahl values and then consider the critical values that are calculated when the simulation only considers observations that create a Herfindahl in the same bin. This conditions the simulation on an inputted Herfindahl range. The larger the bin, the more conservative the critical values will be for a given Herfindahl value in the sense that false positives are avoided. The most conservative critical values will be when the bin is the entire range of Herfindahls, which is the method described above.

For practical application, our program asks the user to specify the number of plants in the industry and a range around the Herfindahl value they have in the data. Given the number of plants, the program takes employment draws as we increment σ, yielding over 100,000 constructed Herfindahl values for that N. The program then finds the largest σ that has at least a 5% chance of generating any Herfindahl value in the range specified by the user. Next the program re-simulates using the inputted N and this largest plausible σ for the specified Herfindahl range. In the re-simulation, the program discards those observations whose calculated Herfindahl is outside of the bin until 10,000 observations that fall within the bin are reached. A γ is calculated from each of those observations in the simulated data and the middle 95% are collected to construct the critical values. This constitutes a critical value that is conditioned on the given Herfindahl bin.

In the appendix we include a table (C.2) that illustrates the output from the program when the Herfindahl range is divided into ten bins. As with table C.1, the results in this table are meant as an illustration of our program output. Table C.2 shows how conditioning on a subrange of Herfindahl values creates critical values given industry competitiveness.

4.2 Geographic Weights

Our program works by first inputting a separate vector of geographic weights $\mathbf{x}$. In our simulations, we let those weights be the state share of non-farm employment from the data. Those weights could be modified to account for observable natural advantages, as in Ellison and Glaeser (1997) and (1999) or for use in local applications.

Result 4. *Increasing the variance of the size of the underlying units of geography increases the width of the confidence interval.*

In addition to our simulations where the geographic weights are the state share of non-farm employment, we also simulated confidence intervals where the geographic weights are drawn from a Dirichlet distribution and a χ^2 distribution. From each distribution, we inputted 850 random $\mathbf{x}$ vectors of length 50 into our program and then ran the simulation as before.

Using a Dirichlet distribution for geographic weights is one way to model natural advantages as in Ellison and Glaeser (1997, p. 900), where the distribution's shape is a function of the natural advantage parameter γ_{na}. We simulate by fixing $\gamma_{na} = 0.1$, which given the values of γ in the data may be large. We take the mean 95% critical values from these 850 geographic weight draws. Also,

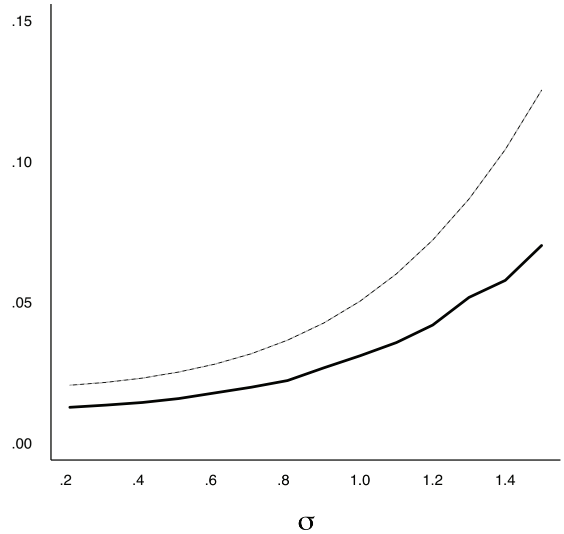


Figure 4. Given a fixed number of plants ($N = 100$), the mean width of the confidence interval to capture 95% mass of 850 geography draws from a Dirichlet distribution with $\gamma_{na} = 0.1$ (dashed line), a χ^2 distribution with $\gamma_{na} = 0.1$ (thin line), and the employment-weighted geography with $\gamma_{na} = 0.0$ (thick line) increases with the standard deviation of the logarithm of the plant employment distribution.

we perform this exercise using a χ^2 distribution for the 850 geographic weight draws. The results are shown in figure 4. The dashed line is the mean width of the confidence internal to capture 95% mass from the 850 Dirichlet draws and the thin line is the mean 95% confidence interval width from the χ^2. The thick line in the figure is the benchmark 95% confidence interval width from the non-farm employment-weighted geography where $\gamma_{na} = 0.0$ and is repeated from figure 1.

Figure 4 shows that the width of the confidence interval to capture 95% mass from the $\gamma_{na} = 0.1$ draws are larger than for the $\gamma_{na} = 0.0$ benchmark regardless of the standard deviation of the plant employment distribution. There is little difference between the mean 95% critical value derived from the simulations using the Dirichlet and χ^2 distributions for geography: the standard errors are larger than the benchmark. That the critical values are larger (in absolute value) is because the Dirichlet and χ^2 distributions result on average in an underlying geography that has both more very large "states" and very small "states" than the distribution of non-farm employment in the data. Thus a large γ could be the result of a normal-sized dart landing in a very small state in addition to a fat dart landing in a normal-sized state as in figure 1.

Geographic weights are important for calculating a critical value. In state, county, or other local applications, we suggest using employment weights. However, if the application is for an industry where natural advantage is suspected to be large, then we recommend modifying the geographic weights to explicitly account for the observed natural advantage such as in Ellison and Glaeser (1999). Inputting those weights into our program results in confidence intervals that are centered around a γ that has accounted for observable natural advantage and thus a statistically significant level of localization would be beyond that expected from observed natural advantage.

5 Which Industries Are Truly Localized?

Using the same 1987 Census of Manufactures data as Ellison and Glaeser (1997), we calculate γ for each of the 459 4-digit SIC manufacturing industries in 1987. The 1987 Census of Manufactures only reports the total industrial employment, number of plants in each of ten employment categories, and the total number of employees in those ten categories except when censoring occurs.[3] It does not report employment in any state-industry with fewer than 150 employees and it reports state-industry employment in categories of 100–249, 250–499, 500–999, 1000–2500, and 2500 plus. Ellison and Glaeser describe the method they use to fill in the unreported data (pgs. 921–5). To estimate the plant Herfindahl for each industry, Ellison and Glaeser use the Schmalensee (1977) method and we use their estimates.[4] (See Feser (2000) and Ellison and Glaeser (1997, pgs. 925–6) for evidence that the Schmalensee method for estimating a Herfindahl matches the data well.)

Ellison and Glaeser (1997) find that all but thirteen industries have $\gamma > 0$, or about 97%, and therefore clustering beyond what is expected from darts thrown on the map is widespread. But since Ellison and Glaeser do not calculate critical values, they do not know how likely it is that a particular observation may have $\gamma > 0$ from randomness alone. They only know $\mathbb{E}[\gamma] = 0$ under the assumption of no natural advantages or within-state spillovers.

The working paper version of Ellison and Glaeser (1994) lists all manufacturing industries, along with their estimated plant Herfindahls. Using our simulated confidence intervals, we are able to perform a statistical test as a function of the Herfindahl to see which industries are statistically

[3] The employment categories for number of plants in each industry and state are 1–4, 5–9, 10–19, 20–49, 50–99, 100–249, 250–499, 500–999, 1000–2499, and 2500 plus.

[4] The 1987 Census of Manufactures did report a firm Herfindahl. We are tremendously grateful to Glenn Ellison who gave us the Ellison and Glaeser (1997) estimates for the unreported data and plant Herfindahls.

localized. Ellison and Glaeser relied on an *ad hoc* threshold of $\gamma > 0.05$ as very localized and $0.02 < \gamma \leq 0.05$ as somewhat localized.

Our results for all 459 manufacturing industries are in appendix C. We use the most conservative 95% upper and lower critical values in statistical testing.

Fact 1. *There are industries with large γ values whose level of localization is not statistically significant.*

We find that 2 of the 127 industries that Ellison and Glaeser deemed very localized have levels of localization that are not different from randomness at a statistically significant level.[5] These are Cellulosic Manmade Fibers (SIC 2823) with $\gamma = 0.159$ and Chewing Gum (SIC 2067) with $\gamma = 0.073$. Cellulosic Fibers has 10,500 employees in 7 plants for a Herfindahl of 0.224 whereas Chewing Gum has 5200 employees in 13 plants for a Herfindahl of 0.157. Thus we attribute the lack of statistical significance to the "fat dart" issue: it is not rare for only 7 or 13 darts to randomly land near each other and have it look like localization because each dart represents many employees. As result 2 shows, when the number of plants is near 10, there is a somewhat large chance of a type I error at the .05 threshold. Since very few national industries in the United States have fewer than 15 plants, it is more of a surprise that there exist any type I errors than that there are just a few of them.

We also find 12 of the 131 industries that Ellison and Glaeser call somewhat localized are not statistically significant. These are listed in table 1. These twelve industries are harder to understand why they are not statistically significant in terms of our simulation results. The number of plants for this group averages 70, employment averages 12,900, and the Herfindahl averages 0.084. We suspect these industries have a large σ, though certainly each of these industries has many fewer plants than the median industry. In section 4 we estimated the chance of a type I error at 2.6%. In the data we find that 14 of 459 industries were misclassified using the Ellison and Glaeser threshold of 0.02, or 3.0%. The rule of thumb seems to be that if there are fewer than 150 plants, there is reason to be concerned for type I error when applying the Ellison and Glaeser thresholds.

[5] In the 1994 working paper, Ellison and Glaeser say they find 119 very localized industries (those with $\gamma > .05$). But using exactly the same data we count 127 very localized industries. Also they report 206 not very localized industries whereas we count 201. We are not sure why this discrepancy exists.

Table 1. Misclassification of Industry Localization in Ellison and Glaeser (1997)

SIC	Name	Employment (thousands)	Plant Herfindahl	Plants	γ
	All Industries With $\gamma > .02$ That Are Not Statistically Significant At 5% Level				
2823	Cellulosic manmade fibers	10.5	.224	7	.159
2067	Chewing gum	5.2	.157	13	.073
2076	Vegetable oil mills, n.e.c	0.9	.084	23	.049
3632	Household refrigerators and freezers	25.7	.107	49	.034
3355	Aluminum rolling and drawing, n.e.c.	0.9	.084	29	.032
3639	Household appliance., n.e.c.	16.0	.061	75	.030
3631	Household cooking equipment	21.9	.050	78	.030
2068	Salted and roasted nuts and seeds	8.8	.079	88	.025
2384	Robes and dressing gowns	8.7	.029	96	.024
3253	Ceramic wall and floor tile	9.5	.039	114	.023
3795	Tanks and tank components	16.7	.157	56	.023
3511	Turbines and turbine generator sets	22.9	.091	81	.023
3463	Nonferrous forgings	7.3	.082	79	.022
3647	Vehicular lighting equipment	15.5	.139	72	.022
	Select Industries With $\gamma < .02$ That Are Statistically Significant At 5% Level				
2711	Newspapers	434.4	.002	9091	.002
2761	Manifold business forms	53.3	.003	856	.002
3444	Sheet metal work	100.2	.001	4296	.003
2026	Fluid Milk	72.4	.002	946	.003
3442	Metal doors, sash, and trim	74.7	.003	1592	.003
2541	Wood partitions and fixtures	40.6	.002	1867	.003
3271	Concrete block and brick	18.6	.002	1128	.004
3086	Plastics foam products	61.3	.004	946	.004
2759	Commercial printing. n.e.c.	125.8	.001	10795	.004
3496	Miscellaneous lubricated wire products	35.1	.003	1157	.004
3569	General industrial machinery. n.e.c.	40.6	.004	1219	.004
3089	Plastics products. n.e.c.	384.9	.001	8571	.005
3953	Marking devices	7.5	.007	636	.005
3446	Architectural metal work	28.0	.004	1345	.005
3082	Unsupported plastics profile shapes	25.2	.007	581	.005

Source: Author's calculations using data described in Ellison and Glaeser (1997).
Note: Only 15 of 112 industries are listed in the bottom half of the table.

Fact 2. *There are many industries with low γ values whose level of localization is statistically significant*

Our simulations show that 112 of 201 industries that Ellison and Glaeser call "not very localized" have levels of localization that are statistically significant, meaning that in fewer than 5% of our simulations did an industry with the same number of plants and employment generate a γ at least as large as in the data. We list the 15 industries with the lowest γ whose levels of localization are statistically significant in table 1. We attribute the statistically significant levels of localization of these industries, despite their low γ values, to the large number of plants. Our simulations show that for a realistic plant employment distribution, once an industry gets to 500 plants, the width of the γ confidence interval is zero to five decimals. For industries having more than the median number of plants, the width of the confidence interval is zero to three decimals for $\sigma < 1$. Because we use the most conservative critical values, switching to less conservative critical values would only add to this list of false negatives.

In section 4 we estimated the chance of a type II error to be 10.6%, but the misclassification in the data occurred for 24.4% of industries. This is because our estimate for the chance of type II was based of fewer than 500 plants. That there are many type II errors is a combination of the result that half of the industries with more than 300 plants have a very narrow confidence interval and that industries with many plants tend to have small plant Herfindahls driving down the calculated γ. This makes type II errors inevitable if a discriminating constant threshold is applied across industries that vary in the number of plants. Though Ellison and Glaeser found 97% of industries had $\gamma > 0$, they said 56% of industries were somewhat or largely localized. Using a 5% level of statistical significance and the most conservative critical values, we find that 78% of industries are localized.

Fact 3. *Diffuse industries exist.*

Ellison and Glaeser find thirteen industries with $\gamma < 0$. We find none of these have levels of localization that are statistically significant at the 5% level. However, in a more recent and larger survey of industrial localization, Holmes and Stevens (2004) calculate the Ellison-Glaeser index for all 1,082 6-digit 1997 NAICS industries using 1999 County Business Patterns data. They find the median γ is 0.020 and the mean is 0.041. While the levels of localization for the most concentrated industries (mostly mining) are all statistically significant, we find that some of their least localized industries also have levels of localization that are statistically significant. In table 2, we list the fifteen least concentrated industries from Holmes and Stevens and indicate those whose level of localization is significant at the 5% level. Those industries whose level of localization is statistically significant can be considered more diffuse than randomness is likely to generate.

What is interesting about the industries that are diffuse is that, other than radio networks (NAICS 515111), they do not have more than 100 plants. However the number of plants cannot be very large for diffuse industries because if there were many plants, they would not be able to spread out enough to be different from darts on the map. Thus each of these industries either has

Table 2. Least Concentrated Industries in Holmes and Stevens (2004)

97 NAICS	Name	Plant Herfindahl	Plants	γ	95% Sig
312213	Engineered wood member (exc truss) mfg	.376	8	-.203	*
485119	Other urban transit systems	.365	27	-.138	*
332995	Other ordnance & accessories mfg	.230	65	-.044	*
521110	Monetary authorities - central bank	.059	46	-.041	*
311312	Cane sugar refining	.110	19	-.040	
325221	Cellulosic organic fiber mfg	.279	10	-.026	
336391	Motor vehicle air-conditioning mfg	.176	70	-.026	
316212	House slipper mfg	.204	20	-.026	
331422	Copper wire (except mechanical) drawing	.062	67	-.021	*
325920	Explosives mfg	.055	95	-.019	
515111	Radio networks	.127	339	-.010	*
325192	Cyclic crude & intermediate mfg	.063	57	-.009	
333397	Scale & balance (except laboratory) mfg	.034	119	-.009	
325413	In-vitro diagnostic substance mfg	.101	223	-.009	
322225	Laminated aluminum foil mfg for flexible pkg	.058	47	-.008	

Source: Author's calculations using data described in Holmes and Stevens (2004).
Note: Industries that have levels of diffusion that are statistically significant at the 5% level are indicated with a $*$.

very similarly sized plants or a relatively wide confidence interval.

6 Conclusion

Ellison and Glaeser (1997) show that a small number of plants may make an industry appear localized when it is not. Their eponymous index γ corrects for this small numbers randomness. They prove that under no natural advantages or spillovers, the expected value of their index is zero. Positive values indicate localization of the industry. But Ellison and Glaeser resorted to *ad hoc* thresholds for deciding if any particular industry is not very localized, somewhat localized, or very localized.

We improve the quantitative interpretation of the Ellison-Gleaser index by simulating confidence intervals that can be used to asses how likely the levels of localization in the data occur from chance alone. We run 100,000 simulations for each combination of two parameters that determine the Ellison-Gleaser index: the number of plants in the industry and standard deviation of the logarithm of the plant employment distribution. We calculate confidence intervals by ordering the 100,000 simulated γ values then selecting the appropriate level of type I error (e.g. 5%) from the top and bottom of our generated distribution and recording the critical values. We change one of the parameters and run another 100,000 simulations.

Our findings show that the width of the confidence interval increases in the standard deviation of the logarithm of the plant employment distribution and decreases with the number of plants in the industry. These findings imply that a constant threshold for determining an industry's level of localization is subject to type I and type II errors. As an illustrative exercise, we use our calculated critical values on all 459 manufacturing industries in the United States in 1987. We find that localization is common: about 78% of manufacturing industries have a level of localization that is statistically significant at a 5% level. However, we find that 2 of Ellison and Glaeser's "very localized" industries and 12 of their "somewhat localized" industries could come from randomness more than 5% of the time. We also find that many of their "not very localized" industries are statistically significant at the 5% level using our most conservative critical values. When we apply our critical values to Holmes and Stevens's (2004) least concentrated industries, we find six industries whose Ellison-Glaeser index is negative but statistically significant, meaning these industries are non-randomly diffuse.

Our results do not indicate whether industries with a statistically significant γ are localized. Rather our results indicate that the same level of localization could be the result of a random placement of plants with given employment more than 5% of the time. In the sense that a researcher is interested in industrial localization beyond that of randomness, then the statistically insignificant industries may not qualify as truly localized. When considering industries at the national level, high plant count industries are the norm resulting in critical values that are dramatically below the *ad hoc* thresholds established by Ellison and Glaeser. This results in a large number of industries where the absolute level of localization is small while still being statistically significant. However, applying any *ad hoc* threshold will result in a trade-off between a relatively large chance of a type II error when applied at the national level and a relatively large chance of a type I error when applied at the local level.

We provide the results of our full simulation in an online appendix at `http://goo.gl/Ox7YD`. This table can be used by a researcher studying localization of any industry at the national level. However, for applications in which the geographic weights need to be changed to account for local conditions or observed natural advantages, then the practitioner should instead input the specific weights into our program and simulate the appropriate confidence intervals. We designed the software such that it is easy to run a simulation under any specification and the desired conservatism.

When interpreting an Ellison-Glaeser index value, one should be careful to see if it is statistically significant. Resorting to a comparison of γ values from other industries, thought to be localized, can be flawed because industries, whose number of plants or standard deviation of the logarithm of the employment distribution differ, can have different γ critical values. We acknowledge that the critical values we report assume the accuracy of the data as our simulations do not account for either poor quality data or that the spillover function is likely to decay over physical distance regardless of regional boundaries. Nevertheless, our simulated confidence intervals provide quantitative meaning to the Ellison-Glaeser index without requiring the heavy data requirements of the Duranton and Overman (2005) index.

References

Bassham, L. E., 2010. A statistical test suite for random and pseudorandom number generators for cryptographic applications. Tech. Rep. 1–131, National Institute of Standards and Technology.

Billings, S. B., Johnson, E. B., Dec. 2013. Agglomeration within an urban area, unpublished.

Briant, A., Combes, P.-P., Lafourcade, M., May 2010. Dots to Boxes: Do the Size and Shape of Spatial Units Jeopardize Economic Geography Estimations? Journal of Urban Economics 67 (3), 287–302.

Cabral, L. M. B., Mata, J., Sep. 2003. On the evolution of the firms size distribution: Facts and theory. American Economic Review 93 (4), 1075–1090.

Combes, P.-P., 2000. Economic Structure and Local Growth: France, 1984–1993. Journal of Urban Economics 47 (3), 329–355.

Deltas, G., 2003. The small-sample bias of the Gini coefficient: Results and implications for empirical research. Review of Economics and Statistics 85, 226–234.

Duranton, G., Overman, H. G., 2005. Testing for localisation using micro-geographic data. Review of Economic Studies 72 (4), 1077–1106.

Ellison, G., Glaeser, E. L., August 1994. Geographic concentration in U.S. manufacturing industries: A dartboard approach, NBER working paper no. 4840, `www.nber.org/papers/w4840.pdf`.

Ellison, G., Glaeser, E. L., Oct. 1997. Geographic concentration in U.S. manufacturing industries: A dartboard approach. Journal of Political Economy 105 (5), 889–927.

Ellison, G., Glaeser, E. L., 1999. The geographic concentration of industry: does natural advantage explain agglomeration? American Economic Review 89 (2), 311–316.

Ferguson, N., Schneier, B., 2003. Practical Cryptography. Wiley Publishing, Inc., New York.

Feser, E. J., June 2000. On the Ellison-Glaeser geographic concentration index, unpublished, www.works.bepress.com/edwardfeser/28.

Gautier, P. A., Teulings, C. N., 2003. An empirical index for labor market density. Review of Economics and Statistics 85 (4), 901–908.

Holmes, T. J., Stevens, J. J., 2004. Spatial distribution of economic activities in North America. In: Henderson, J. V., Thisse, J.-F. (Eds.), Handbook of Urban and Regional Economics. Vol. 4. North Holland, Amsterdam, Ch. 63, pp. 2797–2843.

Kerr, W. R., Kominers, S. D., Dec. 2010. Agglomerative forces and cluster shapes, NBER working paper no. 16639, www.nber.org/papers/w16639.pdf.

Matsumoto, M., Nishimura, T., January 1998. Mersenne twister: A 623-dimensionally equidistant uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation 8 (1), 3–30.

Maurel, F., Sédillot, B., September 1999. A measure of the geographic concentration in French manufacturing industries. Regional Science and Urban Economics 29 (5), 575–604.

Overman, H. G., Puga, D., 2010. Labor pooling as a source of agglomeration: An empirical investigation. In: Glaeser, E. L. (Ed.), Agglomeration Economics. NBER Chapters. National Bureau of Economic Research, Inc, Cambridge, MA, Ch. 7981, pp. 133–150.

Rosenthal, S. S., Strange, W. C., 2001. The Determinants of Agglomeration. Journal of Urban Economics 50 (2), 191–229.

Sawilowsky, S. S., 2003. You think you've got trivials? Journal of Modern Applied Statistical Methods 2 (1), 218–225.

Schmalensee, R., May 1977. Using the H-index of concentration with published data. Review of Economics and Statistics 59 (2), 186–193.

Stanley, M. H. R., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., Stanley, H. E., October 1995. Zipf plots and the size distribution of firms. Economics Letters 49 (4), 453–457.

Appendices

A The Pseudorandom Number Generator

Computers cannot generate truly random numbers. For this simulation, we need a pseudorandom number generator that will not create a pattern in two dimensions. The most common pseudorandom number generator is the Mersenne "Twister" (Matsumoto and Nishimura 1998). When you call a random function in many applications, this is likely the underlying algorithm. Twister is a good algorithm meeting the standards set by Sawilowsky (2003) for Monte Carlo simulations in that it 1) is fast, 2) is unbiased, and 3) has a long repeat cycle. But Twister fails some tests proposed by Bassham (2010) for true randomness. The left panel of figure 5 shows Twister output. It shows the nonrandom pattern of points and emptiness seen as horizontal lines of alternating black and white. Therefore, if we use Twister to generate plant employment and then throw these plant sizes as darts on the map, we would create upwardly biased confidence intervals because the simulation would not think there is clustering when in fact there is a clear pattern.

Instead of Twister, we use the Fortuna pseudorandom number generator. Ferguson and Schneier (2003) show Fortuna meets our requirements preventing random numbers from bunching too much while still being unbiased. The right panel of figure 5 shows Fortuna output. As can be seen, there is no pattern in the black dots and white spaces. The downside of Fortuna is speed. Twister is nearly 150× faster than Fortuna.

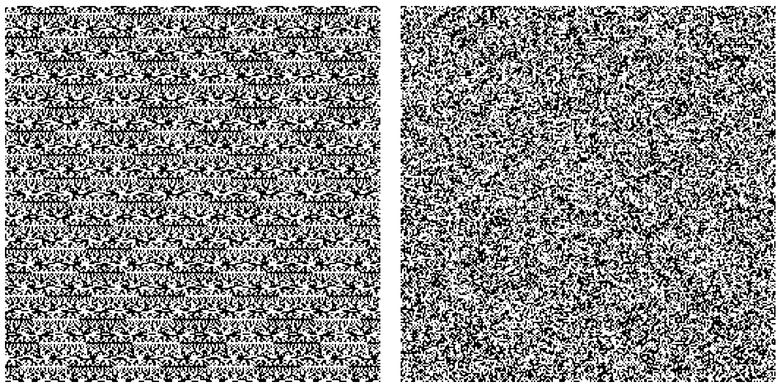


Figure 5. Twister (left panel) versus Fortuna (right panel). Twister produces bunches in two dimension whereas Fortuna does not. This figure is produced with an unrealistically low "k" value, however it better illustrates the clustering nature of the algorithm.

B The Program

Our software requires nothing more than a Unix (including Mac OS X) or Linux system. It can run on Windows, but it requires installing Python. In additional to providing our public domain code at `http://goo.gl/n1N06`, we make it easy to install the program for Mac users because the code is included in the MacPorts repository, `http://www.macports.org/`. With MacPorts installed, one need only type:

```
sudo port -v selfupdate
sudo port install EGSimulation
```

This will automatically download the latest version as well as all dependancies and automatic updating. Once the software is installed, determining the available commands to change the simulation specification is as easy as typing:

```
EGSimulation --help
```

The software allows the practitioner to perform a statistical test on their calculated γ for their specific application. Examples of this would be industry parameters that are not explicitly included in our table or a different $\mathbf{x}$ vector for local applications or to account for observed natural advantages as in Ellison and Glaeser (1999).

C Tables

Table C.1. Simulated Critical Values Using Full Herfindahl Range

Plants	σ	95% Herfindahl Range		5% γ Critical Value	95% γ Critical Value
20	0.20	.051	.053	-.020	.028
20	0.40	.053	.066	-.023	.031
20	0.60	.058	.094	-.026	.036
20	0.70	.060	.116	-.028	.039
20	0.80	.063	.144	-.031	.043
20	0.90	.066	.180	-.034	.047
20	0.95	.067	.199	-.036	.050
20	1.00	.069	.223	-.038	.052
20	1.05	.071	.247	-.040	.055
20	1.10	.073	.270	-.042	.056
20	1.25	.078	.354	-.049	.066
20	1.50	.088	.497	-.062	.081
50	0.20	.021	.021	-.008	.011
50	0.40	.022	.025	-.009	.012
50	0.60	.024	.036	-.011	.015
50	0.70	.026	.044	-.012	.016
50	0.80	.028	.057	-.014	.018
50	0.90	.030	.074	-.015	.020
50	0.95	.031	.083	-.016	.022
50	1.00	.032	.095	-.018	.023
50	1.05	.033	.108	-.019	.024
50	1.10	.034	.124	-.020	.026
50	1.25	.038	.176	-.024	.031
50	1.50	.046	.297	-.033	.042
70	0.20	.015	.015	-.006	.008
70	0.40	.016	.018	-.007	.009
70	0.60	.018	.025	-.008	.010
70	0.70	.019	.031	-.009	.012
70	0.80	.021	.039	-.010	.013
70	0.90	.022	.051	-.011	.015
70	0.95	.023	.059	-.012	.016
70	1.00	.024	.068	-.013	.017
70	1.05	.025	.078	-.014	.018
70	1.10	.026	.090	-.015	.020
70	1.25	.029	.133	-.018	.024
70	1.50	.036	.235	-.026	.033
100	0.20	.010	.011	-.004	.005
100	0.40	.011	.012	-.005	.006
100	0.60	.013	.017	-.006	.007
100	0.70	.014	.021	-.006	.008
100	0.80	.015	.027	-.007	.009
100	0.90	.016	.035	-.008	.011
100	0.95	.017	.040	-.009	.012
100	1.00	.018	.046	-.009	.012
100	1.05	.018	.054	-.010	.013
100	1.10	.019	.062	-.011	.014
100	1.25	.022	.097	-.014	.018
100	1.50	.028	.183	-.020	.026
150	0.20	.007	.007	-.003	.004
150	0.40	.008	.008	-.003	.004
150	0.60	.009	.011	-.004	.005
150	0.70	.009	.013	-.004	.006
150	0.80	.010	.017	-.005	.006
150	0.90	.011	.023	-.006	.007
150	0.95	.012	.026	-.006	.008
150	1.00	.012	.030	-.006	.008
150	1.05	.013	.035	-.007	.009
150	1.10	.014	.041	-.008	.010
150	1.25	.016	.066	-.010	.013
150	1.50	.020	.132	-.015	.019
200	0.20	.005	.005	-.002	.003
200	0.40	.006	.006	-.002	.003
200	0.60	.007	.008	-.003	.004
200	0.70	.007	.010	-.003	.004
200	0.80	.008	.013	-.004	.005
200	0.90	.009	.016	-.004	.006
200	0.95	.009	.019	-.005	.006
200	1.00	.010	.022	-.005	.007
200	1.05	.010	.026	-.005	.007
200	1.10	.011	.030	-.006	.008
200	1.25	.013	.048	-.008	.010
200	1.50	.016	.103	-.012	.015
250	0.20	.004	.004	-.002	.002
250	0.40	.005	.005	-.002	.002
250	0.60	.005	.006	-.002	.003
250	0.70	.006	.008	-.003	.003
250	0.80	.006	.010	-.003	.004
250	0.90	.007	.013	-.003	.004
250	0.95	.007	.015	-.004	.005
250	1.00	.008	.017	-.004	.005
250	1.05	.008	.020	-.004	.006
250	1.10	.009	.024	-.005	.006
250	1.25	.010	.039	-.006	.008

Continued on next page

Table C.1 – *Continued from previous page*

Plants	σ	95% Herfindahl Range		5% γ Critical Value	95% γ Critical Value
250	1.50	.014	.086	-.010	.012

Source: Author's calculations.

The Ellison-Glaeser index is γ and σ is the standard deviation of the logarithm of the plant employment distribution. To use this table, first find the number of plants to match the data. Next, scan over the 95% Herfindahl ranges within that number of plants and settle on the Herfindahl ranges that match the data. The critical values are the lower and upper bound for which 95% of random observations lie between. In order for γ to be statistically significant, the value must be outside of this range. Since Herfindahl ranges are not unique, the most conservative critical values would be the widest range of γ critical values, which could span multiple rows. Find the complete table at `http://goo.gl/Ox7YD`.

Table C.2. Simulated Critical Values Using Conditional Herfindahl Bins

Plants	Herfindahl Bin		max σ	5% γ Critical Value	95% γ Critical Value
20	.0516	.0545	0.4	-.0231	.0330
20	.0545	.0590	0.6	-.0249	.0334
20	.0590	.0651	0.8	-.0274	.0384
20	.0651	.0728	1.0	-.0297	.0404
20	.0728	.0821	1.2	-.0334	.0457
20	.0821	.0934	1.4	-.0368	.0502
20	.0934	.1076	1.5	-.0423	.0601
20	.1076	.1272	1.5	-.0468	.0699
20	.1272	.1622	1.5	-.0556	.0841
20	.1622	.2924	1.5	-.0767	.1041
50	.0207	.0220	0.3	-.0096	.0129
50	.0220	.0241	0.5	-.0102	.0138
50	.0241	.0271	0.7	-.0113	.0155
50	.0271	.0310	0.9	-.0129	.0175
50	.0310	.0359	1.0	-.0147	.0192
50	.0359	.0421	1.2	-.0166	.0224
50	.0421	.0498	1.5	-.0197	.0265
50	.0498	.0604	1.5	-.0234	.0313
50	.0604	.0783	1.5	-.0280	.0365
50	.0783	.1485	1.5	-.0381	.0524
70	.0148	.0157	0.3	-.0068	.0093
70	.0157	.0173	0.5	-.0073	.0103
70	.0173	.0195	0.7	-.0083	.0112
70	.0195	.0225	0.8	-.0093	.0123
70	.0225	.0264	1.0	-.0108	.0143
70	.0264	.0312	1.2	-.0123	.0167
70	.0312	.0372	1.4	-.0147	.0197
70	.0372	.0452	1.5	-.0177	.0236
70	.0452	.0589	1.5	-.0212	.0283
70	.0589	.1125	1.5	-.0294	.0397
100	.0104	.0110	0.3	-.0047	.0064
100	.0110	.0121	0.5	-.0052	.0068
100	.0121	.0137	0.6	-.0057	.0079
100	.0137	.0159	0.8	-.0066	.0091
100	.0159	.0188	1.0	-.0076	.0100
100	.0188	.0226	1.2	-.0092	.0124
100	.0226	.0273	1.4	-.0109	.0142
100	.0273	.0334	1.5	-.0128	.0176
100	.0334	.0438	1.5	-.0160	.0209
100	.0438	.0831	1.5	-.0221	.0285
150	.0069	.0073	0.3	-.0031	.0044
150	.0073	.0081	0.5	-.0034	.0048
150	.0081	.0092	0.6	-.0038	.0054
150	.0092	.0108	0.8	-.0045	.0060
150	.0108	.0128	1.0	-.0053	.0071
150	.0128	.0155	1.1	-.0062	.0083
150	.0155	.0190	1.3	-.0075	.0104
150	.0190	.0235	1.5	-.0092	.0127
150	.0235	.0308	1.5	-.0112	.0152
150	.0308	.0577	1.5	-.0161	.0218
250	.0042	.0044	0.3	-.0019	.0025
250	.0044	.0049	0.4	-.0020	.0027
250	.0049	.0056	0.6	-.0023	.0032
250	.0056	.0065	0.8	-.0028	.0038
250	.0065	.0079	0.9	-.0032	.0045
250	.0079	.0096	1.1	-.0039	.0054
250	.0096	.0120	1.3	-.0048	.0063
250	.0120	.0149	1.5	-.0060	.0078
250	.0149	.0196	1.5	-.0073	.0097
250	.0196	.0368	1.5	-.0105	.0140

Source: Author's calculations.

The Ellison-Glaeser index is γ and σ is the standard deviation of the logarithm of the plant employment distribution. To use this table, first find the row with the correct number of plants. Next, find the appropriate bin for your Herfindahl. The third column is the largest σ that has at least a 5% chance of generating any value in that bin. The critical values are the lower and upper bound for which 95% of random observations lie between conditional on those observations having a Herfindahl value inside that bin and with that number of plants. In order for γ to be statistically significant, the value must be outside of this range.

Table C.3. Reproduction of Ellison and Glaeser SIC 4 with Significance at 5% Level

SIC	Industry	Employment (thousands)	Plant Herfindahl	Plants	γ	EG Localized	95% Sig
2011	Meat packing plants	113.9	.008	1434	.042	Y	*
2013	Sausages and other prepared meats	78.7	.004	1343	.006		*
2015	Poultry slaughtering and Processing	147.9	.005	463	.054	YY	*
2021	Creamery butter	1.7	.045	49	.147	YY	*
2022	Cheese, natural and processed	33.0	.009	644	.131	YY	*
2023	Dry, condensed and evaporated dairy products	14.1	.056	186	.015		
2024	Ice Cream & Frozen Desserts	20.3	.008	541	.000		
2026	Fluid Milk	72.4	.002	946	.003		*
2032	Canned Specialities	24.5	.032	211	-.012		
2033	Canned, Fruits and Vegetables	65.1	.006	647	.044	Y	*
2034	Dehydrated fruits, vegetables and soups	10.1	.030	132	.280	YY	*
2035	Pickles, sauces and salad dressings	21.4	.013	382	-.001		
2037	Frozen fruits and vegetables	49.8	.011	258	.079	YY	*
2038	Frozen specialities n.e.c	37.5	.015	288	.002		
2041	Flour and other grain mill products	13.3	.009	358	.018		*
2043	Cereal breakfast foods	16.0	.054	53	.018		
2044	Rice milling	4.5	.053	63	.136	YY	*
2045	Prepared flour mixes and doughs	12.1	.020	149	.014		
2046	Wet corn milling	8.6	.050	60	.138	YY	*
2047	Dog and cat food	13.4	.018	186	.011		*
2048	Prepared feeds, n.e.c	34.5	.002	1738	.019		*
2051	Bread, cake and related products	161.9	.003	2357	.000		
2052	Cookies and crackers	45.3	.028	379	-.001		
2053	Frozen bakery products except bread	9.9	.035	114	.013		
2061	Raw cane sugar	6.2	.038	40	.289	YY	*
2062	Cane sugar refining	5.5	.107	21	.000		
2063	Beet sugar	7.9	.031	42	.074	YY	*
2064	Candy and other confectionary products	45.8	.012	685	.046	Y	*
2066	Chocolate and cocoa products	11.0	.107	186	.038	Y	*
2067	Chewing gum	5.2	.157	13	.073	YY	
2068	Salted and roasted nuts and seeds	8.8	.079	88	.025	Y	
2074	Cottonseed oil mills	2.6	.032	52	.168	YY	*
2075	Soybean oil mills	7.0	.020	106	.070	YY	*
2076	Vegetable oil mills, n.e.c	.9	.084	23	.049	Y	
2077	Animal and marine fats and oils	9.8	.009	305	.011		*
2079	Edible fats and oils, n.e.c	9.3	.021	100	.031	Y	*
2082	Malt beverages	31.9	.042	134	-.010		
2083	Malt	1.4	.072	27	.238	YY	*
2084	Wines, brandy and brandy spirits	13.9	.041	508	.479	YY	*
2085	Distilled and blended liquors	9.0	.035	72	.079	YY	*
2086	Bottled and canned soft drinks	95.6	.002	1190	.005		*
2087	Flavoring extracts and syrups n.e.c	9.1	.018	280	.025	Y	*
2091	Canned and cured fish and seafoods	6.7	.020	175	.061	YY	*
2092	Fresh or frozen prepared fish	38.2	.007	645	.059	YY	*
2095	Roasted coffee	10.7	.026	141	.032	Y	*
2096	Potato chips and similar snacks	33.1	.011	344	.009		*
2097	Manufactured Ice	4.7	.006	549	.011		*
2098	Macaroni and spaghetti	6.6	.028	218	-.001		
2099	Food preparations, n.e.c	58.0	.003	1658	.014		*
2111	Cigarettes	32.0	.223	12	.169	YY	*
2121	Cigars	2.5	.107	20	.158	YY	*
2131	Chewing and smoking tobacco	3.3	.083	29	.200	YY	*
2141	Tobacco stemming and redrying	6.9	.045	76	.177	YY	*
2211	Broadwoven fabric mills, cotton	72.3	.025	301	.170	YY	*
2221	Broadwoven fabric mills, manmade fiber and silk	88.3	.007	436	.228	YY	*
2231	Broadwoven fabric mills, wool	14.0	.042	118	.087	YY	*
2241	Narrow fabric mills	18.5	.011	272	.074	YY	*
2251	Women's hosiery, except socks	29.3	.028	161	.398	YY	*
2252	Hosiery, n.e.c	36.5	.008	426	.437	YY	*
2253	Knit outerwear mills	59.0	.012	824	.065	YY	*
2254	Knit underwear mills	19.3	.082	63	.019		
2257	Weft knit fabric mills	34.9	.019	334	.191	YY	*
2258	Lace and warp knit fabric mills	20.5	.014	240	.116	YY	*
2259	Knitting mills, n.e.c	3.8	.071	79	.094	YY	*
2261	Finishing plants, cotton	16.5	.019	198	.124	YY	*
2262	Finishing plants, manmade	27.9	.022	268	.188	YY	*
2269	Finishing plants, n.e.c	11.7	.020	182	.098	YY	*
2273	Carpets and rugs	53.3	.013	475	.378	YY	*
2281	Yarn spinning mills	89.0	.005	414	.284	YY	*
2282	Throwing and winding mills	18.3	.025	139	.206	YY	*
2284	Thread mills	6.5	.051	59	.207	YY	*
2295	Coated fabrics, not rubberized	10.3	.020	185	.000		

Continued on next page

Table C.3 – *Continued from previous page*

SIC	Industry	Employment (thousands)	Plant Herfindahl	Plants	γ	EG Localized	95% Sig
2296	Tire cord and fabrics	5.1	.121	13	.178	YY	*
2297	Nonwoven fabrics	13.8	.023	130	.039	Y	*
2298	Cordage and twine	6.9	.017	197	.033	Y	*
2299	Textile goods, n.e.c	16.4	.009	551	.021	Y	*
2311	Men's and boys' suits and coats	55.2	.010	337	.042	Y	*
2321	Men's and boys' shirts	76.7	.004	601	.062	YY	*
2322	Men's and boys' underwear and nightwear	17.2	.032	96	.096	YY	*
2323	Men's and boys' neckwear	7.4	.018	142	.106	YY	*
2325	Men's and boys' trousers and slacks	93.3	.004	484	.064	YY	*
2326	Men's and boys' work clothing	33.1	.009	255	.090	YY	*
2329	Men's and boys' clothing, n.e.c	52.2	.006	616	.025	Y	*
2331	Women's, misses', and juniors' blouses and shirts	73.4	.002	1496	.038	Y	*
2335	Women's, misses', and juniors' dresses	112.7	.001	5471	.098	YY	*
2337	Women's, misses', and juniors' suits and coats	55.2	.003	1092	.034	Y	*
2339	Women's, misses', and juniors' outerwear, n.e.c	107.3	.002	2198	.028	Y	*
2341	Women's and children's underwear	53.7	.006	434	.053	YY	*
2342	Brassieres, girdles and allied garments	13.8	.024	128	.019		
2353	Hats, caps and millnery	17.2	.013	462	.044	Y	*
2361	Girls' and children's dresses and blouses	30.9	.007	454	.030	Y	*
2369	Girls' and children's outerwear, n.e.c	40.8	.008	381	.046	Y	*
2371	Fur goods	2.2	.007	380	.630	YY	*
2381	Fabric dress and work gloves	4.8	.027	82	.103	YY	*
2384	Robes and dressing gowns	8.7	.029	96	.024	Y	
2385	Waterproof outerwear	6.4	.057	67	.075	YY	*
2386	Leather and sheep-lined clothing	2.1	.034	131	.100	YY	*
2387	Apparel belts	10.5	.013	265	.167	YY	*
2389	Apparel and accessories, n.e.c	8.3	.015	340	.020	Y	*
2391	Curtains and draperies	27.1	.008	1250	.025	Y	*
2392	Housefurnishings n.e.c	50.5	.006	944	.036	Y	*
2393	Textile bags	8.8	.011	262	.005		
2394	Canvas and related products	16.7	.005	1274	.010		*
2395	Pleating and stitching	14.1	.009	685	.026	Y	*
2396	Automotive and apparel trimmings	44.2	.016	1558	.074	YY	*
2397	Schiffli machine embroideries	5.9	.025	271	.153	YY	*
2399	Fabricated textile products, n.e.c	30.5	.008	916	.005		*
2411	Logging	85.8	.001	11937	.061	YY	*
2421	Sawmills and planing mills, general	148.3	.001	5741	.038	Y	*
2426	Hardwood dimension and flooring mills	29.7	.005	737	.063	YY	*
2429	Special product sawmills, n.e.c	2.2	.009	234	.374	YY	*
2431	Millwork	89.0	.005	2783	.013		*
2434	Wood kitchen cabinets	67.0	.002	3714	.011		*
2435	Hardwood veneer and plywood	20.5	.008	311	.050	Y	*
2436	Softwood veneer and plywood	38.9	.008	232	.187	YY	*
2439	Structural wood members, n.e.c	24.6	.003	893	.026	Y	*
2441	Nailed wood boxes and shook	5.9	.009	308	.018		*
2448	Wood pallets and skids	25.7	.001	1701	.006		*
2449	Wood containers, n.e.c	5.4	.023	208	.026	Y	*
2451	Mobile homes	39.9	.005	395	.037	Y	*
2452	Prefabricated wood buildings	25.4	.006	689	.024	Y	*
2491	Wood preserving	11.8	.005	540	.028	Y	*
2493	Reconstituted wood products	22.0	.011	240	.028	Y	*
2499	Wood products, n.e.c	56.3	.002	3324	.006		*
2511	Wood household furniture	135.9	.003	2949	.077	YY	*
2512	Upholstered household furniture	82.1	.004	1150	.131	YY	*
2514	Metal household furniture	30.1	.010	418	.013		*
2515	Mattresses and bedsprings	24.4	.004	839	.007		*
2517	Wood television and radio cabinets	5.9	.072	81	.010		
2519	Household furniture, n.e.c.	5.9	.050	177	.004		
2521	Wood office furniture	31.0	.009	649	.045	Y	*
2522	Office furniture, except wood	49.7	.036	337	.050	Y	*
2531	Public building and related furniture	21.8	.012	491	.008		
2541	Wood partitions and fixtures	40.6	.002	1867	.003		*
2542	Parlitions and fixtures, except wood	33.5	.007	592	.010		*
2591	Drapery hardware and blinds and shades	20.6	.018	489	.006		
2599	Furniture and fixtures, n.e.c.	29.3	.005	1597	.007		*
2611	Pulp mills	14.2	.051	39	.047	Y	*
2621	Paper mills	129.1	.008	282	.039	Y	*
2631	Paperboard mills	52.3	.011	205	.024	Y	*
2652	Setup paperboard boxes	8.7	.011	200	.037	Y	*
2653	Corrugated and solid fiber boxes	105.7	.001	1600	.001		
2655	Fiber cans, drums, and similar products	12.5	.009	281	.006		*
2656	Sanitary food container	15.8	.047	92	.028	Y	*
2657	Folding paperboard boxes	50.7	.004	606	.002		
2671	Papercoated and laminated packaging	15.0	.018	120	.018		*
2672	Paper coated and laminated, n.e.c.	30.9	.017	412	.010		*
2673	Bags: plastics, laminated, and coated	36.6	.009	483	.011		*
2674	Bags: uncoated paper and multiwall	17.1	.013	132	.025	Y	*
2675	Die-cut paper and board	15.7	.011	399	.010		*
2676	Sanitary paper products	38.4	.020	133	.033	Y	*
2677	Envelopes	27.6	.007	298	.008		*
2678	Stationery products	11.2	.021	189	.024	Y	*
2679	Converted paper products, n.e.c.	29.6	.009	821	.011		*
2711	Newspapers	434.4	.002	9091	.002		*

Continued on next page

Table C.3 – *Continued from previous page*

SIC	Industry	Employment (thousands)	Plant Herfindahl	Plants	γ	EG Localized	95% Sig
2721	Periodicals	110.0	.005	4020	.067	YY	*
2731	Book publishing	70.1	.008	2298	.062	YY	*
2732	Book printing	43.5	.012	561	.011		*
2741	Miscellaneous publishing	69.5	.005	2369	.008		*
2752	Commercial printing, lithographic	403.9	.000	24984	.004		
2754	Commercial printing. gravure	23.8	.032	332	.017		*
2759	Commercial printing. n.e.c.	125.8	.001	10795	.004		*
2761	Manifold business forms	53.3	.003	856	.002		*
2771	Greeting cards	21.5	.091	162	.037	Y	*
2782	Blankbooks and looseleaf binders	39.1	.007	510	.008		*
2789	Bookbinding and related work	29.7	.005	1036	.020		*
2791	Typesetting	37.6	.002	3364	.015		*
2796	Platemaking services	31.8	.002	1413	.010		*
2812	Alkalies and chlorine	5.0	.061	45	.058	YY	*
2813	Industrial gases	8.1	.005	594	.011		*
2816	Inorganic pigments	8.3	.041	92	.031	Y	*
2819	Industrial inorganic chemicals, n.e.c.	72.2	.053	662	.017		*
2821	Plastics materials and resins	56.3	.012	480	.029	Y	*
2822	Synthetic rubber	10.4	.063	68	.164	YY	*
2823	Cellulosic manmade fibers	10.5	.224	7	.159	YY	
2824	Organic fibers. noncellulosic	45.4	.043	71	.140	YY	*
2833	Medicinals and botanicals	11.6	.042	225	.088	YY	*
2834	Pharmaceutical preparations	131.6	.015	732	.023	Y	*
2835	Diagnostic substances	15.4	.033	158	.059	YY	*
2836	Biological products, except diagnostic	13.3	.023	241	.010		
2841	Soap and other detergents	31.7	.016	764	.003		
2842	Polishes and sanitation goods	20.6	.010	726	.018		*
2843	Surface active agents	9.1	.017	217	.040	Y	*
2844	Toilet preparations	57.9	.011	694	.054	YY	*
2851	Paints and allied products	55.2	.003	1428	.007		*
2861	Gum and wood chemicals	2.6	.041	77	.061	YY	*
2865	Cyclic crudes and intermediates	22.8	.019	186	.009		
2869	Industrial organic chemicals. n.e.c.	100.3	.012	699	.069	YY	*
2873	Nitrogenous fertilizers	7.4	.025	164	.031	Y	*
2874	Phosphatic fertilizers	9.4	.066	77	.290	YY	*
2875	Fertilizers, mixing only	7.5	.006	452	.020	Y	*
2879	Agricultural chemicals. n.e.c.	16.1	.038	277	.031	Y	*
2891	Adhesives and sealants	20.9	.005	714	.012		*
2892	Explosives	13.8	.113	132	.003		
2893	Printing Ink	11.1	.005	504	.015		*
2895	Carbon black	1.8	.054	22	.300	YY	*
2899	Chemical preparations. n.e.c.	37.9	.006	1531	.005		*
2911	Petroleum refining	74.6	.011	308	.089	YY	*
2951	Asphalt paving mixtures and blocks	14.6	.003	1101	.009		*
2952	Asphalt felt, and coatings	13.5	.009	266	.010		*
2992	LubrIcating oils and greases	11.2	.007	451	.013		*
2999	Petroleum and coal products, n.e.c.	1.9	.027	106	.062	YY	*
3011	Tires and Inner tubes	65.4	.025	163	.038	Y	*
3021	Rubber and plastics footwear	10.9	.060	65	-.013		
3052	Rubber and plastics hose and belting	23.2	.026	188	.038	Y	*
3053	Gaskets, packing, and sealing devices	28.4	.011	496	.015		*
3061	Mechanical rubber goods	49.8	.008	624	.047	Y	*
3069	Fabricated rubber products, n.e.c.	54.3	.006	1009	.023	Y	*
3081	Unsupported plastics film and sheet	48.4	.006	594	.007		*
3082	Unsupported plastics profile shapes	25.2	.007	581	.005		*
3083	Laminated plastics plate, sheet, and profile shapes	17.3	.025	234	.005		
3084	Plastics pipe	12.5	.008	251	.010		*
3085	Plastics bottles	25.1	.007	286	.012		*
3086	Plastics foam products	61.3	.004	946	.004		*
3087	Custom compounding of purchased plastics resins	17.3	.008	405	.012		*
3088	Plastics plumbing fixtures	7.5	.023	176	.015		
3089	Plastics products. n.e.c.	384.9	.001	8571	.005		*
3111	Leather tanning and finishing	14.6	.013	344	.025	Y	*
3131	Footwear cut stock	5.0	.032	127	.141	YY	*
3142	House slippers	3.7	.104	37	.066	YY	*
3143	Men's footwear, except athletic	31.6	.016	154	.073	YY	*
3144	Womens footwear, except athletic	26.6	.012	163	.055	YY	*
3149	Footwear, except rubber, n.e.c.	9.2	.025	129	.087	YY	*
3151	Leather gloves and mittens	3.1	.028	77	.034	Y	*
3161	Luggage	11.4	.027	241	.042	Y	*
3171	Women's handbags and purses	9.5	.021	321	.144	YY	*
3172	Personal leather goods. n.e.c.	7.2	.024	209	.059	YY	*
3199	Leather goods, n.e.c.	7.1	.011	396	.024	Y	*
3211	Flat glass	14.6	.055	84	.019		
3221	Glass containers	41.1	.013	106	.011		*
3229	Pressed and blown glass. n.e.c.	36.3	.020	416	.038	Y	*
3231	Products of purchased glass	51.1	.005	1429	.002		
3241	Cement, hydraulic	19.1	.009	213	.010		*
3251	Brick and structural clay tile	16.6	.007	266	.036	Y	*
3253	Ceramic wall and floor tile	9.5	.039	114	.023	Y	
3255	Clay refractories	6.4	.027	153	.078	YY	*
3259	Structural clay products. n.e.c.	2.1	.048	67	.160	YY	*
3261	Vitreous plumbing fixtures	9.7	.041	65	.014		

Continued on next page

Table C.3 – *Continued from previous page*

SIC	Industry	Employment (thousands)	Plant Herfindahl	Plants	γ	EG Localized	95% Sig
3262	vitreous china table and kitchenware	5.4	.126	34	.000		
3263	Semivitreous table and kitchenware	1.8	.109	44	.088	YY	*
3264	Porcelain electrical supplies	10.7	.030	116	.045	Y	*
3269	Pottery products, n.e.c.	10.5	.016	754	.012		*
3271	Concrete block and brick	18.6	.002	1128	.004		*
3272	Concrete products. n.e.c.	70.0	.001	3154	.012		*
3273	Readymixed concrete	96.8	.001	5319	.010		*
3274	Lime	5.7	.033	82	.064	YY	*
3275	Gypsum products	12.1	.013	152	.013		*
3281	Cut stone and stone products	12.5	.011	746	.036	Y	*
3291	Abrasive products	23.4	.038	405	.028	Y	*
3292	Asbestos products	4.0	.107	54	.008		
3295	Minerals, ground or treated	8.8	.011	381	.006		
3296	Mineral wool	21.5	.020	231	.015		*
3297	Nonclay retractories	7.7	.020	135	.043	Y	*
3299	NonmetallIc mineral products, n.e.c.	7.6	.009	543	.004		
3312	Blast furnaces and steel mills	188.1	.018	342	.068	YY	*
3313	Electrometallurgial products	3.9	.072	30	.148	YY	*
3315	Steel wire and related products	24.7	.012	343	.013		*
3316	Cold finishing of steel shapes	16.4	.027	191	.032	Y	*
3317	Steel pipe and tubes	19.6	.010	221	.038	Y	*
3321	Gray and ductile iron foundries	82.4	.011	774	.028	Y	*
3322	Malleable iron foundries	4.2	.197	28	.072	YY	*
3324	Steel investment foundries	20.3	.040	135	-.003		
3325	Steel foundries, n.e.c.	22.9	.012	294	.040	Y	*
3331	Primary copper	3.3	.135	13	.194	YY	*
3334	Primary aluminum	17.3	.050	49	.053	YY	*
3339	Primary nonferrous metals. n.e.c.	11.0	.044	108	.005		
3341	Secondary nonferrous metals	12.5	.008	398	.015		*
3351	Copper rolling and drawing	22.6	.029	121	.017		
3353	Aluminum sheet, plate, and foil	26.1	.063	56	.009		
3354	Aluminum extruded products	30.7	.013	204	.001		
3355	Aluminum rolling and drawing, n.e.c.	.9	.084	29	.032	Y	
3356	Nonferrous rolling and drawing. n.e.c.	17.9	.031	172	.016		
3357	Nonferrous wiredrawing and insulating	64.9	.008	487	.017		*
3363	Aluminum die-castings	28.1	.010	412	.021	Y	*
3364	Nonferrous die-casting, except aluminum	12.9	.010	304	.036	Y	*
3365	Aluminum foundries	26.3	.008	583	.021	Y	*
3366	Copper foundries	8.2	.007	334	.013		*
3369	Nonferrous foundries, n.e.c.	4.0	.117	56	.103	YY	*
3398	Metal heat treating	18.0	.004	725	.026	Y	*
3399	Primary metal products. n.e.c.	13.8	.105	252	.060	YY	*
3411	Metal cans	39.4	.006	369	.009		*
3412	Metal barrels, drums, and pails	8.7	.014	168	.042	Y	*
3421	Cutlery	10.5	.039	141	.056	YY	*
3423	Hand and edge tools. n.e.c.	41.9	.008	810	.008		*
3425	Saw blades and handsaws	7.7	.039	138	.039	Y	*
3429	Hardware, n.e.c.	85.2	.007	1239	.009		*
3431	Metal sanitary ware	8.0	.064	97	.030	Y	*
3432	Plumbing fixture fittings and trim	17.1	.023	180	.003		
3433	Heating equipment, except electric	20.5	.008	556	.001		
3441	Fabricated structural metal	80.9	.006	2453	.004		
3442	Metal doors, sash, and trim	74.7	.003	1592	.003		*
3443	Fabricated plate work (boiler shops)	74.7	.004	1740	.010		*
3444	Sheet metal work	100.2	.001	4296	.003		*
3446	Architectural metal work	28.0	.004	1345	.005		*
3448	Prefabricated metal buildings	25.8	.009	560	.006		
3449	Miscellaneous metal work	22.9	.006	597	.015		*
3451	Screw machine products	42.7	.002	1635	.027	Y	*
3452	Bolts, nuts, rivets, and washers	52.0	.006	937	.029	Y	*
3462	Iron and steel forgings	26.6	.017	406	.024	Y	*
3463	Nonferrous forgings	7.3	.082	79	.022	Y	
3465	Automotive stampings	119.8	.013	713	.177	YY	*
3466	Crowns and closures	6.1	.056	57	.039	Y	*
3469	Metal stampings. n.e.c.	95.5	.002	2815	.017		*
3471	Plating and polishing	71.1	.001	3451	.013		*
3479	Metal coating and allied services	41.5	.002	1814	.015		*
3482	Small arms ammunition	9.0	.184	79	-.004		
3483	Ammunition, except tot small arms, n.e.c.	41.5	.041	87	.003		
3484	Small arms	13.3	.067	151	.080	YY	*
3489	Ordnance and accessories. n.e.c.	23.9	.166	59	.004		
3491	Industrial valves	45.9	.009	384	.006		
3492	Fluid power valves and hose fittings	27.9	.010	386	.038	Y	*
3493	Steel springs, except wire	5.0	.024	151	.048	Y	*
3494	Valves and pipe fittings, n.e.c.	25.1	.010	416	.017		*
3495	Wire springs	19.7	.009	407	.014		*
3496	Miscellaneous labricated wire products	35.1	.003	1157	.004		*
3497	Metal foil and leaf	10.4	.033	117	.033	Y	*
3498	Fabricated pipe arid fittings	20.0	.004	728	.021	Y	*
3499	Fabricated metal products. n.e.c.	72.5	.002	3782	.006		*
3511	Turbines and turbine generator sets	22.9	.091	81	.023	Y	
3519	Internal combustion engines, n.e.c.	64.0	.034	278	.070	YY	*
3523	Farm machinery and equipment	57.0	.013	1634	.063	YY	*

Continued on next page

Table C.3 – *Continued from previous page*

SIC	Industry	Employment (thousands)	Plant Herfindahl	Plants	γ	EG Localized	95% Sig
3524	Lawn and garden equipment	24.9	.043	165	.014		
3531	Construction machinery	81.1	.016	954	.060	YY	*
3532	Mining machinery	13.6	.016	321	.057	YY	*
3533	Oil and gas field machinery	24.8	.015	633	.433	YY	*
3534	Elevators and moving stairways	10.2	.028	176	-.002		
3535	Conveyors and conveying equipment	31.5	.005	747	.018		*
3536	Hoists, cranes, and monorails	7.0	.020	175	.015		
3537	Industrial trucks and tractors	20.1	.016	467	.004		
3541	Machine tools, metal cutting types	31.7	.019	417	.035	Y	*
3542	Machine tools, metal forming types	13.8	.018	207	.071	YY	*
3543	Industrial patterns	8.6	.006	813	.051	YY	*
3544	Special dies, tools, jigs, and fixtures	114.4	.001	7317	.053	YY	*
3545	Machine tool accessories	48.5	.003	1881	.037	Y	*
3546	Power-driven handtools	16.8	.037	199	.045	Y	*
3547	Rolling mill machInery	3.9	.067	86	.085	YY	*
3548	Welding apparatus	18.7	.028	225	.040	Y	*
3549	Metalworking machinery. n.e.c.	11.3	.011	301	.040	Y	*
3552	Textile machinery	15.6	.012	506	.165	YY	*
3553	Woodworking machinery	8.9	.016	292	.033	Y	*
3554	Paper industries machinery	17.1	.022	278	.096	YY	*
3555	Printing trades machinery	25.0	.032	438	.017		*
3556	Food products machinery	19.2	.008	512	.015		*
3559	Special industry machinery. n.e.c.	83.3	.003	2531	.007		*
3561	Pumps and pumping equipment	35.2	.010	405	.009		
3562	BaIl and roller bearings	36.9	.021	169	.043	Y	*
3563	Air and gas compressors	23.8	.021	259	.020	Y	*
3564	Blowers and fans	24.8	.008	507	.003		
3565	Packaging machinery	22.6	.010	439	.018		*
3566	Speed changers, drives, and gears	17.9	.019	276	.019		*
3567	Industrial furnaces and ovens	16.6	.010	370	.005		
3568	Power transmission equipment. n.e.c.	22.0	.014	308	.014		*
3569	General Industrial machinery. n.e.c.	40.6	.004	1219	.004		*
3571	Electronic conputers	151.9	.019	974	.058	YY	*
3572	Computer storage devices	43.3	.113	106	.142	YY	*
3575	Computer terminals	15.0	.046	121	.004		
3577	Computer peripheral equipment, n.e.c.	76.2	.030	549	.031	Y	*
3578	Calculating and accounting equipment	12.8	.060	98	.009		
3579	Office machines. n.e.c.	28.5	.053	204	.015		
3581	Automatic vending machines	7.9	.062	98	.004		
3582	Commercial laundry equipment	4.6	.054	81	.020		
3585	Refrigeration and heating equipment	133.3	.008	894	.011		*
3586	Measuring and dispensing pumps	9.4	.083	83	.002		
3589	Service Industry machinery, n.e.c.	35.2	.005	949	.014		*
3592	Carburetors, pistons, rings, and valves	21.7	.038	155	.042	Y	*
3593	Fluid power cylinders and actuators	20.2	.052	362	.026	Y	*
3594	Fluid power pumps and motors	14.8	.034	150	.002		
3596	Scales and balances, except laboratory	6.7	.027	134	.023	Y	*
3599	Industrial machinery. n.e.c.	228.5	.000	21547	.005		
3612	Trartsformers, except electronic	32.2	.016	286	.021	Y	*
3613	Switchgear and switchboard apparatus	44.8	.010	474	.008		
3621	Motors and generators	74.6	.008	462	.022	Y	*
3624	Carbon and graphite products	9.8	.033	95	.042	Y	*
3625	Relays and industrial controls	66.6	.010	1168	.008		*
3629	Electrical industrial apparatus. n.e.c.	14.5	.017	481	.010		*
3631	Household cooking equipment	21.9	.050	78	.030	Y	
3632	Household reirigerators and freezers	25.7	.107	49	.034	Y	
3633	Household laundry equipment	16.7	.128	18	.124	YY	*
3634	Electric housewares and fans	25.1	.019	230	.107	YY	*
3635	Household vacuum cleaners	11.3	.182	31	-.008		
3639	Household appliance., n.e.c.	16.0	.061	75	.030	Y	
3641	Electric lamp bulbs and tubes	22.2	.027	127	.032	Y	*
3643	Current-carrying wiring devices	47.9	.017	430	.009		*
3644	Noncurrent-carrying wiring devices	21.5	.023	209	.011		
3645	Residential lighting fixtures	22.5	.009	580	.027	Y	*
3646	Commercial lighting fixtures	22.7	.022	271	.019		*
3647	Vehicular lighting equipment	15.5	.139	72	.022	Y	
3648	Lighting equipment, n.e.c.	14.4	.017	262	.011		
3651	Household audio and video equipment	30.9	.035	378	.016		*
3652	Prerecorded records and tapes	13.3	.039	476	-.008		
3661	Telephone and telegraph apparatus	112.3	.021	469	.009		*
3663	Radio and television communications equipment	126.0	.015	655	.020	Y	*
3669	Communications equipment, n.e.c.	21.9	.017	382	.030	Y	*
3671	Electron tubes	28.4	.057	121	.043	Y	*
3672	Printed circuit boards	66.6	.005	1009	.041	Y	*
3674	Semiconductors and related devices	184.6	.014	853	.064	YY	*
3675	Electronic capacitors	21.7	.023	148	.029	Y	*
3676	Electronic resistors	15.7	.022	118	.016		*
3677	Electronic coils and transformers	23.9	.009	416	.018		*
3678	Electronic connectors	42.8	.017	271	.035	Y	*
3679	Electronic components, n.e.c.	162.6	.008	2900	.023	Y	*
3691	Storage batteries	24.2	.017	190	.010		
3692	Primary batteries, dry and wet	10.7	.045	72	.049	Y	*
3694	Engine electrical equipment	67.3	.045	487	.054	YY	*

Continued on next page

Table C.3 – *Continued from previous page*

SIC	Industry	Employment (thousands)	Plant Herfindahl	Plants	γ	EG Localized	95% Sig
3695	Magnetic and optical recording media	25.6	.028	200	.084	YY	*
3699	Electrical equipment and supplies, n.e.c.	60.3	.008	1379	.015		*
3711	Motor vehicIes and car bodies	281.3	.016	413	.127	YY	*
3713	Truck and bus bodies	37.8	.009	716	.008		*
3714	Motor vehicle parts and accessories	389.6	.006	2807	.089	YY	*
3715	Truck trailers	27.5	.013	337	.014		*
3716	Motor homes	15.1	.055	165	.149	YY	*
3721	Aircraft	268.2	.053	155	.023	Y	*
3724	Aircraft engines and engine parts	139.6	.042	453	.046	Y	*
3728	Aircraft parts and equipment n.e.c.	188.2	.029	1014	.031	Y	*
3731	Ship building and repairing	120.2	.080	590	.014		*
3732	Boat building and repairing	57.2	.005	2176	.046	Y	*
3743	Railroad equipment	22.1	.085	174	.123	YY	*
3751	Motorcycles, bicycles, and parts	7.4	.077	246	.010		
3761	Guided missiles and space vehicles	166.7	.046	40	.249	YY	*
3764	Space propulsion units and parts	31.8	.145	35	.111	YY	*
3769	Space vehicle equipment. n.e.c.	15.1	.157	66	.004		
3792	Travel trailers and campers	17.2	.011	427	.087	YY	*
3795	Tanks and tank components	16.7	.157	56	.023	Y	
3799	Transportation equipment. n.e.c.	15.4	.015	635	.021	Y	*
3812	Search and navigation equipment	369.4	.011	1084	.040	Y	*
3821	Laboratory apparatus and furniture	17.1	.020	260	-.001		
3822	Environmental controls	26.5	.035	254	.011		
3823	Process control instruments	53.3	.010	784	.017		*
3824	Fluid meters and counting devices	10.1	.032	158	.022	Y	*
3825	Instruments to measure electricity	85.2	.014	930	.031	Y	*
3826	Analytical instruments	31.2	.014	562	.039	Y	*
3827	Optical instruments and lenses	20.1	.027	250	.061	YY	*
3829	Measuring and contolling devices, n.e.c.	41.0	.015	970	.004		
3841	Surgical and medical instruments	73.1	.007	1136	.011		*
3842	Surgical appliances and supplies	78.5	.005	1501	.005		*
3843	Dental equipment and supplies	14.6	.017	505	.023	Y	*
3844	X-ray apparatus and tubes	8.7	.049	75	.017		
3845	Electromedical equipment	29.2	.021	224	.025	Y	*
3851	Opthalmic goods	24.2	.020	495	.027	Y	*
3861	Photographic equipment and supplies	88.0	.067	787	.174	YY	*
3873	Watches, clocks, watchcases, and parts	11.8	.031	218	.005		
3911	Jewelry, precious metal	35.5	.005	2324	.094	YY	*
3914	Silverware and plated ware	6.9	.065	209	.049	Y	*
3915	Jewelers' materials and lapidary work	7.1	.025	442	.298	YY	*
3931	Musical instruments	12.2	.017	423	.015		*
3942	Dolls and stuffed toys	4.4	.027	197	.086	YY	*
3944	Games, toys, and childrens vehicles	30.9	.017	716	.011		*
3949	Sporting and athletic goods, n.e.c.	53.6	.005	1800	.003		
3951	Pens and mechanical pencils	8.4	.048	110	.030	Y	*
3952	Lead pencils and art goods	5.6	.045	145	.030	Y	*
3953	Marking devices	7.5	.007	636	.005		*
3955	Carbon paper and inked ribbons	7.3	.035	125	.008		
3961	Costume jewelry	22.2	.017	760	.320	YY	*
3965	Fasteners, buttons, needles, and pins	9.6	.018	262	.041	Y	*
3991	Brooms and brushes	12.3	.014	301	.006		
3993	Signs and advertising specialties	66.3	.001	3778	.006		*
3995	Burial caskets	8.7	.026	231	.050	YY	*
3996	Hard surface floor coverings, n.e.c.	7.6	.139	21	.097	YY	*
3999	Manufacturing Industries, n.e.c.	68.3	.003	4093	.008		*

Source: Author's calculations using data described in Ellison and Glaeser (1997). A single "Y" in the EG localized column indicates a γ value above 0.02 while "YY" indicates above 0.05. A "*" in the "95% Sig" column indicates that the industry is localized beyond randomness using the most conservative critical values.