Firms and Industry Agglomeration*

Dominick Bartelme
University of Michigan

Oren Ziv
Michigan State University

January 2020

Abstract

Industry agglomeration can be indicative of agglomeration forces and is correlated with firm outcomes. Because multi-plant firms tend cluster their establishments in space, industry agglomeration could in part be driven by forces internal to the firm rather than across-firm spillovers. We propose and implement a decomposition of the industry agglomeration measures into within and across-firm components using U.S. census microdata. The within-firm component makes a small contribution to observed industry agglomeration for most industries and spatial scales, but accounts for 20% or more of observed agglomeration at short spatial scales for a subset of industries.

JEL codes: R32,L25

Keywords: economic geography, agglomeration, firm location, firm growth

*We thank Andrei Levchenko for comments and suggestions, and C. Luke Watson for excellent research assistance. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Contact: dbartelm@umich.edu, orenziv@msu.edu.

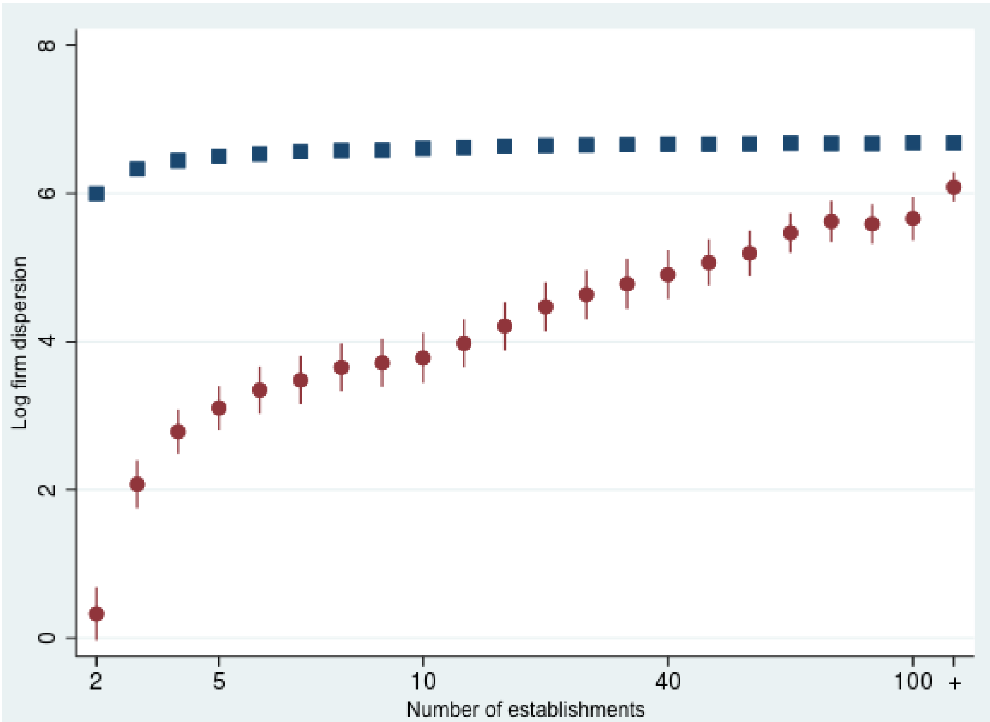
1 Introduction

Measures of industry agglomeration are crucial spatial statistics, in part because they suggest the potential strength of agglomeration forces that generate them. Recent studies on the spatial structure of firms finds that firms cluster their constituent establishments more closely than the typical group of establishments in that industry (Behrens and Sharunova, 2015; Bartelme and Ziv, 2020), as shown in Figure 1. This finding raises the possibility that measured industry agglomeration levels may reflect within-firm establishment clustering as well as industry-level clustering. Existing measures do not account for the role of within-firm clustering in industry agglomeration (e.g. Ellison and Glaeser (1997); Duranton and Overman (2005)). Because the forces governing within-firm plant location decisions are likely different from those governing plant location decisions across firms, it is important to quantify the extent to which observed industry agglomeration is due to within-firm vs. across-firm location decisions. This distinction becomes especially important when forces such as productivity or knowledge spillovers are invoked to explain spatial agglomerations and to draw policy conclusions, because optimal policy depends on whether these forces are internal or external to firms.

This paper proposes and implements a decomposition of the (employment-weighted) density of bilateral plant distances introduced by Duranton and Overman (2005) into within and across-firm components. Differences between total industry agglomeration and across-firm agglomeration at any spatial scale can be attributed to a combination of a) differences between the within and across firm distributions at that scale, and b) the relative weight of each distribution in the overall density. Using this decomposition, we propose two measures to quantify the contribution of within-firm agglomeration to the overall level of agglomeration in an industry at different spatial scales. We then implement this decomposition using confidential US Census microdata on the location and firm structure of manufacturing establishments in 2012, with industries measured at the 4-digit NAICS level.

We find that on average the within-firm density is much higher than the across-firm density at short spatial scales, in line with the findings in Behrens and Sharunova (2015) and Bartelme and Ziv (2020). For example, in the median industry two employees se-

Figure 1: Firm clustering relative to industry



Note: Red circles plot average distance between plants and firm centroids in each establishment number category. Blue squares plot the same measure for a synthetically generated baseline in which within-firm establishment location choices reflect only industry (6-digit NAICS) clustering patterns. From Bartelme and Ziv (2020): see paper for more details.

lected at random from a single firm are 4.9 times more likely to be within 20 miles of each other than two randomly selected employees at different firms, with the difference rising to 15.6. at the 90th percentile. Nevertheless, the overall contribution of the within-firm component to total industry agglomeration is modest (less that 3% at the median) for most industries at most spatial scales, because the overall proportion of within-firm bilateral pairs (which determines the weight on the within-firm component of the distribution) is low for most industries. However, there are some exceptions to this rule: for industries at the 90th percentile, around 20% of bilateral pairs within 20 miles of one another are within-firm. We conclude that within-firm location patterns are usually not important for explaining industry agglomeration, but can significantly affect industry agglomeration measures at short distances in a subset of industries that are dominated by a relatively small number of firms.[1]

Our paper contributes to the sizable literature on measuring industry agglomeration, e.g. (Ellison and Glaeser, 1997; Marcon and Puech, 2003; Duranton and Overman, 2005; Delgado et al., 2015; Buzard et al., 2017; Marcon and Puech, 2017; Kopczewska et al., 2019). Our contribution to this literature is to introduce a new decomposition of industry agglomeration into within and across-firm components and document the relative importance of each in U.S. manufacturing data. Our paper also contributes to the literature that attempts to explain the sources of industry agglomeration (e.g. Rosenthal and Strange (2001); Ellison et al. (2010); Alfaro and Chen (2014); Billings and Johnson (2016)) as well as the correlation between agglomeration and economic outcomes (e.g. Porter (2000); Holmes and Stevens (2002); Henderson (2003); Kerr and Kominers (2015); Alfaro et al. (2019)). The relationship between agglomeration and outcomes is typically conceived as resulting from the causal impact of agglomeration forces that are external to the firm, or from selection. Our paper raises the possibility that the internal geographic structure of multi-unit firms also partly accounts for both the observed levels of industry agglomeration and their correlations with economic outcomes. In practice we find that this channel is unlikely to be a major driver of agglomeration patterns in most industries.

2 Decomposition

Consider a fixed number J plants indexed by i or j at a set of fixed potential locations, and a joint probability distribution of plant locations. Denote the resulting density of non-zero ($i \neq j$) bilateral plant distances by $\tilde{f}(d)$. Duranton and Overman (2005) compare estimates of this density to the counterfactual density that would arise if plants were located randomly in order to identify "excess" agglomeration of plants at different spatial scales. They also consider the density of bilateral distances between workers in different plants, which we denote by $f(d)$, that can be interpreted as an employment-weighted version of the density of plant distances. In the rest of the paper we work

[1] We provide a list of the industries with the largest and smallest contributions of the within-firm component in the Appendix.

with the employment-weighted version, although our theoretical results extend to the unweighted version and we obtain substantially similar empirical results as well.

We denote the presence of firms in the model by dividing the plants into $S \leq J$ mutually exclusive groups, indexed by s, of one or more plants each. We say that the non-zero ($i \neq j$) bilateral distance between pair of workers at plants i and j, with $i \neq j$, is "within-firm" if both $i \in s$ and $j \in s$, and "across-firm" otherwise. We denote the density of within-firm bilateral distances by $f^w(d)$ and the across-firm distances by $f^a(d)$. With these definitions, we can represent the overall density $f(d)$ as a mixture of the within and across densities,

$$f(d) = \alpha f^w(d) + (1 - \alpha) f^a(d), \tag{1}$$

where the mixture weight α is the proportion of bilateral distances that are within-firm.

We use this decomposition to propose several measures of the contribution of the within-firm component to the overall density of bilateral distances $f(d)$. Our first measure is the probability that a randomly chosen pair of workers at distance d from each other belong to the same firm, given by

$$p(W|d) = \frac{\alpha f^w(d)}{f(d)}. \tag{2}$$

This quantity measures the contribution of the within-firm component to the overall density in an accounting sense. Our second measure is the percent difference between the actual density $f(d)$ and the density that would be observed if the within-firm distribution was identical to the across-firm distribution, which is

$$\tilde{p}(d) = \frac{f(d)}{f^a(d)} - 1 = \alpha \left(\frac{f^w(d)}{f^a(d)} - 1 \right). \tag{3}$$

This quantity measures the contribution of the within-firm distribution relative to a particular counterfactual scenario in which firms allocated their plants in space in a manner similar to the industry as a whole. This measure can also be interpreted as a version of (2) that adjusts for the fact that, were the within-firm component to disappear, the across-firm component would have to be adjusted to integrate to 1. While there are other counterfactual scenarios that could be considered, we view this as a natural and simple benchmark.[2] However measured, the contribution of the within-firm component can be quantitatively significant only if a) α is relatively large, and b) $f^w(d)$ is quantitatively different than $f^a(d)$.

3 Implementation

Data and implementation To implement our decomposition, we use restricted-access US Census microdata from the 2012 Longitudinal Business Database. Our sample includes all manufacturing establishments (NAICS 31-33) with at least one employee and a valid zipcode, with industries measured at the 4-digit NAICS level (86 total). Our final sample consists of 11 million manufacturing employees, two-thirds of which are employed at firms with at least two locations.

We discretize the densities by integrating over five distance intervals and estimating the resulting quantities from the data. For an interval $d = (d_1, d_2]$, we compute $p(W|d) \approx \alpha \int_d f^w(x)\,dx / \int_d f(x)\,dx$. We report the distribution of results across industries for each interval by ranking the results by industry and reporting the relevant measure at the 10th, 25th, 50th, 75th, and 90th percentiles.[3]

Results Table 1 reports values of the ratio $f^w(d)/f^a(d)$ for industries at five different percentiles as well as the mean industry's value. The within-firm density is much greater than the across-firm density at short distances. In the median industry, a randomly chosen pair of workers at different plants within the same firm are almost 5 times more likely to be within 20 miles of one another than if they belonged to a different firm. For some industries, the ratio is much larger: for the 90th-percentile industry, workers at dif-

[2] In practice the two measures are quite similar, as shown in the following section. The difference between the two measures can be written as

$$p(W|d) - \tilde{p}(d) = \alpha \cdot \frac{f(d)\,[f^a(d) - f^w(d)] + f^w(d) f^a(d)}{f(d) f^a(d)}.$$

When $f(d) \approx f^a(d)$, as is the case when α is small, then $p(W|d) - \tilde{p}(d) \approx \alpha$. When $f(d) \approx f^w(d)$, which is the case when α is large, then $p(W|d) - \tilde{p}(d) \approx -\alpha f^w(d)/f^a(d)$. When $f^w(d) \approx f^a(d)$ then the two measures are approximately identical. Empirically α tends to be small.

[3] In order to pass Census disclosure avoidance tests, we report, for each percentile measure, the mean of three measures around that percentile. For example, the 90th percentile result is the simple mean of the values for industries with the 8th, 9th, and 10th largest values.

Table 1: Ratio of within to across firm densities, by percentile

Distance	0-20 miles	20-40 miles	40-60 miles	60-160 miles	160-300 miles
10th Percentile	0.9	0.4	0.0	0.4	0.5
25th Percentile	2.8	1.2	0.6	0.9	0.8
50th Percentile	4.9	2.1	1.8	1.4	1.1
75th Percentile	10.4	4.4	3.3	1.9	1.4
90th Percentile	15.6	6.9	5.2	2.9	1.6
Mean	7.1	3.3	3.2	1.9	1.1

Note: Table 1 reports industry-level percentiles for the ratio $f^w(d)/f^a(d)$ using the discretized densities. Intervals are exclusive of the left and inclusive of the right cutoffs. Percentile measures are mean of three industries' values around a given percentile.

Table 2: Counterfactual contribution of within-firm component, by percentile

Distance	0-20 miles	20-40 miles	40-60 miles	60-160 miles	160-300 miles
10th Percentile	0.0	-0.5	-1.7	-1.2	-0.5
25th Percentile	0.9	0.1	-0.4	-0.1	-0.1
50th Percentile	2.9	0.8	0.3	0.2	0.1
75th Percentile	7.2	3.3	2.2	1.0	0.4
90th Percentile	18.4	10.3	5.7	2.7	1.1
Mean	7.2	2.9	2.4	0.5	0.2

Note: Table 2 reports industry-level percentiles for $\tilde{p}(d)$ using the discretized densities. Intervals are exclusive of the left and inclusive of the right cutoffs. Percentile measures are mean of three industries' values around a given percentile.

ferent establishments of the same firm are more than 15 times more likely to be within 20 miles of each other than workers at different firms. Differences narrow for larger distances. Overall, a necessary condition for the relevance of the within-firm component is satisfied: firm employment is much more agglomerated than industry employment, especially at short distances.

Table 2 reports our measure of the counterfactual contribution of the within-firm component to industry clustering, $\tilde{p}(d)$, for the same intervals and percentiles as Table 1. We focus on this measure because both measures yield very similar results.[4] For

[4] Rank correlations between the two measures above 0.95 for most distance intervals. We report the

most industries, the contribution of the within-firm component to spatial agglomeration is small at all spatial scales. For example, for the median industry, the contribution is around 3% at the 0-20 mile range and declines to below 1% at longer distances. However, the contribution is non-negligible at shorter distances for a subset of of industries. For industries at the 90th percentile, two randomly chosen workers that are within 20 miles of one another (but not at the same plant) have a roughly 1 in 5 chance of working at the same firm, and 1 in 10 in the range of 20-40 miles. Thus there are certainly some industries and spatial scales for which observed agglomeration is influenced by within-firm agglomeration to a significant extent.

Which factors make within-firm clustering more relevant to industry agglomeration? Across industries, $\tilde{p}(d)$ is correlated with α (correlation of 0.60 at 0-20 miles) and $f^w(d)$ (correlation of 0.17 at 0-20 miles, 0.46 at 20-40 miles); the within-firm component significantly influences overall agglomeration measures in industries where fewer firms dominate and in industries where within-firm clustering is especially high.[5] On the other hand, $\tilde{p}(d)$ is essentially uncorrelated with total industry agglomeration $f(d)$ (correlation of -0.06 at 0-20 miles) and weakly negatively correlated with industry employment (correlation of -0.24 at 0-20 miles). In the Appendix, we provide a list of the industries that are most and least affected by the within-firm component.

4 Conclusion

Overall, the presence of firm clustering has a moderate impact on the level of industry agglomeration, with the impact being larger at shorter distances. A few industries are significantly impacted, especially those dominated by fewer firms. Studies of agglomeration, coagglomeration and clustering that include these industries should account for the within-firm component.

results for $p(W|d)$ in the Appendix.

[5] A formal variance decomposition of the contribution of each factor to $\tilde{p}(d)$ is not especially informative because it involves both sums and products of correlated random variables.

References

Alfaro, Laura and Maggie Xiaoyang Chen, "The global agglomeration of multinational firms," *Journal of International Economics*, 2014, *94* (2), 263–276.

_ , Maggie X Chen, and Harald Fadinger, "Spatial Agglomeration and Superstar Firms: Firm-level Patterns from Europe and US," 2019.

Bartelme, Dominick and Oren Ziv, "The Internal Geography of Firms," *Working Paper*, 2020.

Behrens, Kristian and Vera Sharunova, "Inter-and intra-firm linkages: Evidence from microgeographic location patterns," *CEPR Discussion Papers*, 2015, (10921).

Billings, Stephen B and Erik B Johnson, "Agglomeration within an urban area," *Journal of Urban Economics*, 2016, *91*, 13–25.

Buzard, Kristy, Gerald A Carlino, Robert M Hunt, Jake K Carr, and Tony E Smith, "The agglomeration of American R&D labs," *Journal of Urban Economics*, 2017, *101*, 14–26.

Delgado, Mercedes, Michael E Porter, and Scott Stern, "Defining clusters of related industries," *Journal of Economic Geography*, 2015, *16* (1), 1–38.

Duranton, Gilles and Henry G Overman, "Testing for localization using micro-geographic data," *The Review of Economic Studies*, 2005, *72* (4), 1077–1106.

Ellison, Glenn and Edward L Glaeser, "Geographic concentration in US manufacturing industries: a dartboard approach," *Journal of Political Economy*, 1997, *105* (5), 889–927.

_ , _ , and William R Kerr, "What causes industry agglomeration? Evidence from coagglomeration patterns," *American Economic Review*, 2010, *100* (3), 1195–1213.

Henderson, J Vernon, "Marshall's scale economies," *Journal of Urban Economics*, 2003, *53* (1), 1–28.

Holmes, Thomas J and John J Stevens, "Geographic concentration and establishment scale," *Review of Economics and Statistics*, 2002, *84* (4), 682–690.

Kerr, William R and Scott Duke Kominers, "Agglomerative forces and cluster shapes," *Review of Economics and Statistics*, 2015, *97* (4), 877–899.

Kopczewska, Katarzyna, Paweł Churski, Artur Ochojski, and Adam Polko, "SPAG: Index of spatial agglomeration," *Papers in Regional Science*, 2019, *98* (6), 2391–2424.

Marcon, Eric and Florence Puech, "Evaluating the geographic concentration of industries using distance-based methods," *Journal of Economic Geography*, 2003, *3* (4), 409–428.

_ and _ , "A typology of distance-based measures of spatial concentration," *Regional Science and Urban Economics*, 2017, *62*, 56–67.

Porter, Michael E, "Location, competition, and economic development: Local clusters in a global economy," *Economic Development Quarterly*, 2000, *14* (1), 15–34.

Rosenthal, Stuart S and William C Strange, "The determinants of agglomeration," *Journal of Urban Economics*, 2001, *50* (2), 191–229.

A Appendix

Table A1: Contribution of within-firm component, by percentile

Distance	0-20 miles	20-40 miles	40-60 miles	60-160 miles	160-300 miles
10th Percentile	0.1	0.2	0.0	0.1	0.1
25th Percentile	1.7	0.5	0.3	0.4	0.4
50th Percentile	3.6	1.9	1.1	1.3	1.0
75th Percentile	7.8	5.6	3.9	2.7	2.1
90th Percentile	17.1	12.5	6.9	4.2	4.4
Mean	7.1	4.0	3.3	1.9	1.7

Note: Table A1 reports industry-level percentiles for $p(W|d)$ using the discretized densities.

Table A2: List of industries with within-firm contribution $>$ 10%

Results awaiting Census disclosure

Note: Table reports the list of industries for which $\tilde{p}(d) > 10\%$ for some d. For all industries, d is either 0-20 or 20-40.

Table A3: List of industries with within-firm contribution $< 1\%$

Results awaiting Census disclosure

Note: Table reports the list of industries for which $\tilde{p}(d) < 1\%$ for all d.