JUE Insight: Firms and Industry Agglomeration*

Dominick Bartelme
University of Michigan

Oren Ziv
Michigan State University

May 2021

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. Declarations of interest: none. Corresponding author: Oren Ziv. Contact: dbartelm@umich.edu, orenziv@msu.edu.

1 Introduction

Measures of industry agglomeration are crucial spatial statistics, in part because they suggest the strength of the 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). This finding raises the possibility that measured industry-level agglomeration may reflect within-firm establishment clustering as well as industry-level clustering of unrelated establishments. 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)).

The distinction between within-firm clustering and across-firm clustering is relevant for understanding both the causes of industry agglomeration and the implications for optimal policy. The literature typically explains industry-level agglomeration by a mix of exogenous local characteristics and endogenous outcomes of concentrated economic activity, such as thick local markets for specialized labor and other inputs, or localized knowledge flows (see Ellison et al., 2010; Rosenthal and Strange, 2004). The strength of these endogenous agglomeration forces may vary depending on whether they occur within or across firms; for example, "thick input markets" explanations presuppose the presence of many firms, while knowledge may flow more easily within firms than across. Hence knowing whether observed agglomeration is within or across firm is informative for explaining its ultimate causes. Furthermore, the optimal policy depends on whether these endogenous agglomeration forces are internal or external to firms. To the extent that observed agglomerations reflect across-firm forces there may be policy-relevant externalities; to the extent that they reflect within-firm agglomeration forces, they are already internalized by the firms when making their location decisions.

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 selected at random from a single firm are about 5 times more likely to be within 20 miles of each other than two randomly selected employees at different firms, with the difference rising to about 17 at the 90th percentile for manufacturing and about 37 for services. Nevertheless, the overall contribution of the within-firm component to total industry agglomeration is modest (around 2% 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, between 16% (manufacturing) and 22% (services) of bilateral pairs within 20 miles of one another are within-firm. We provide a list of the most industries for which the within-firm component is most significant, which includes some classic examples of highly agglomerated industries such as textiles, auto manufacturing and several IT and financial industries. 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.

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)) and a growing literature on the geography of multi-establishment firms (Behrens and Sharunova, 2015; Bartelme and Ziv, 2020; Lan, 2019; Hsieh and Rossi-Hansberg, 2019; Oberfield et al., 2020). 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 with the employment-weighted version, although our theoretical results extend to the unweighted version and the empirical results are similar in the two cases as well.

We incorporate 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 a 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 $f(d)$.[1] Our first measure is 100 $\times$ the probability that a randomly chosen pair of workers at distance d from each other belong to the same firm, given by

$$p(W|d) = 100 \cdot \frac{\alpha f^w(d)}{f(d)}. \tag{2}$$

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

$$\tilde{p}(d) = 100 \cdot \alpha \left(\frac{f^w(d)}{f^a(d)} - 1 \right). \tag{3}$$

This quantity measures the contribution of the within-firm distribution relative to a counterfactual scenario in which firms allocated their plants in space in a manner similar to the industry as a whole (the "random location" benchmark). 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 adjust in order to integrate to 1. While there are other counterfactual scenarios that could be considered, we view this as a natural and simple benchmark. In practice, the two measures are quite

[1] In practice these statistics will be most useful when the empirical density is approximated by binning the data; see Section 3. If convenient, one can think of $f(d)$ as a probability mass function and d as a distance interval when interpreting the formulas below.

similar and we therefore focus on reporting $\tilde{p}(d)$ in the empirical section.[2] However measured, the contribution of the within-firm component can be quantitatively significant only if a) α is relatively large, and b) the ratio $f^w(d)/f^a(d)$ is significantly greater than 1.

3 Implementation

Data and implementation To implement our decomposition, we use restricted-access US Census microdata from the 2012 Longitudinal Business Database (LBD). Our sample includes all establishments with at least one employee and a valid zipcode, and with an establishment-level industry code at the 4-digit NAICS level in either manufacturing (NAICS 31-33) or services (NAICS 42-56 and 71-81). We identify plants as establishments in the LBD.[3]

We use firm identifiers in the LBD to group establishments into firms within each 4-digit NAICS industry.

Some service industries in our sample are large. To deal with computational limitations for those industries, we take 10% random sample of firms in industries with over 25,000 establishments and 1% random sample of firms in industries with over 250,000 establishments. In order to avoid introducing bias, we sample firms within firm size categories so as to preserves the overall firm size distribution in each industry.[4]

We discretize the densities by integrating over five distance intervals and estimate the resulting quantities from the data. For an interval $d = (d_1, d_2]$, we compute

$$f^w(d)/f^a(d) \approx \int_d f^w(x)\,dx \Big/ \int_d f^a(x)\,dx \tag{4}$$

[2] The difference between the two measures can be written as

$$p(W|d) - \tilde{p}(d) = 100 \cdot \alpha \cdot \frac{f(d)\,[f^a(d) - f^w(d)] + f^w(d) f^a(d)}{f(d) f^a(d)}.$$

When $f(d) \approx f^a(d)$, as is the case when α is small, then $p(W|d) - \tilde{p}(d) \approx 100 \cdot \alpha$. When $f(d) \approx f^w(d)$, which is the case when α is large, then $p(W|d) - \tilde{p}(d) \approx -100\alpha f^w(d)/f^a(d)$. When $f^w(d) \approx f^a(d)$, the two measures are approximately identical. Empirically α tends to be small.

[3] Census defines establishments of multi-unit firms as units operating at physically distinct locations (addresses). We discuss alternative ways of defining establishments in the appendix.

[4] To satisfy Census disclosure avoidance criteria, we also drop three 4-digit service industries.

and

$$\tilde{p}(d) \approx 100 \cdot \alpha \left(\int_d f^w(x)\,dx \Big/ \int_d f^a(x)\,dx - 1 \right). \tag{5}$$

We focus attention on short and medium spatial scales by choosing cutoffs of 0, 20, 40, 60, 160, and 300 miles. 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.[5]

Given that we are estimating $f^w(d)/f^a(d)$ and $\tilde{p}(d)$ for a large number of industries and distance intervals, we should expect to observe some industries with large realizations of these statistics even if the underlying distribution is the random location model ($f^w(d)/f^a(d) = 1$). To check this possibility, we generate a random-location benchmark by simulating one hundred economies where plant locations remain as in the data but firm links are randomly re-assigned within industries. For each statistic computed from actual data, we also report the fifth highest realization of that statistic from the 100 random location benchmarks. For example, we compute the 90th percentile value of $\tilde{p}(20)$ in each of the 100 draws and reports the 5th highest result. If the actual value for a given statistic exceeds the upper range of the random location benchmark for the corresponding statistic, it is unlikely that the data was generated by the random location model.

Results Table 1 reports values of the ratio $f^w(d)/f^a(d)$ for manufacturing and service industries at five different percentiles. The observed within-firm density is typically much greater than the across-firm density at short distances. For the median manufacturing or service industry, a randomly chosen pair of workers at different plants within the same firm are about 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 manufacturing industry, workers at different establishments of the same firm are over 15 times more likely to be within 20 miles of each other than workers at different firms, and the difference is even greater for the 90th-percentile services

[5] 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
			(A) Manufactures		
10th Percentile, Observed	0.79	0.27	0.03	0.26	0.50
Upper C.I. of Random Benchmark	(0.01)	(0.01)	(0.01)	(0.40)	(0.58)
25th Percentile	2.22	0.97	0.50	0.75	0.80
Upper C.I. of Random Benchmark	(0.25)	(0.27)	(0.28)	(0.72)	(0.84)
50th Percentile	5.28	1.80	1.43	1.36	1.13
Upper C.I. of Random Benchmark	(0.73)	(0.78)	(0.84)	(1.03)	(1.06)
75th Percentile	10.93	4.42	2.94	1.87	1.44
Upper C.I. of Random Benchmark	(1.39)	(1.38)	(1.53)	(1.37)	(1.36)
90th Percentile	17.50	6.86	5.22	3.11	1.84
Upper C.I. of Random Benchmark	(2.88)	(2.82)	(3.17)	(2.02)	(1.88)
			(B) Services		
10th Percentile, Observed	1.23	0.14	0.22	0.34	0.25
Upper C.I. of Random Benchmark	(0.37)	(0.11)	(0.08)	(0.40)	(0.60)
25th Percentile	2.06	0.83	0.82	0.91	0.77
Upper C.I. of Random Benchmark	(0.78)	(0.58)	(0.55)	(0.80)	(0.87)
50th Percentile	4.57	1.56	1.46	1.30	1.11
Upper C.I. of Random Benchmark	(1.07)	(0.91)	(0.90)	(0.99)	(1.02)
75th Percentile	14.58	4.13	2.83	2.21	1.52
Upper C.I. of Random Benchmark	(1.59)	(1.25)	(1.18)	(1.19)	(1.20)
90th Percentile	37.22	10.68	6.55	5.59	2.51
Upper C.I. of Random Benchmark	(2.72)	(1.88)	(1.81)	(1.58)	(1.58)

Notes: Table 1 reports industry-level percentiles for the ratio $f^w(d)/f^a(d)$ using the discretized densities. Numbers in parentheses are the 5th highest values of the corresponding moments out of 100 counterfactual economies in which plants are randomly assigned to firms without replacement. In order to comply with Census disclosure avoidance requirements all percentile measures are means of three industries' values around a given percentile.

industry. Differences between the two densities narrow for larger distances, although $f^w(d)/f^a(d)$ continues to be greater than 1 for the median industry at the 160-300 mile range. These differences consistently exceed the extreme values of the random benchmark simulations for distance intervals less than 160 miles, indicating that the observed differences are likely to be systematic. 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

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
			(A) Manufactures		
10th Percentile, Observed	-0.52	-0.45	-1.54	-1.18	-0.37
Upper C.I. of Random Benchmark	(-0.58)	(-0.61)	(-0.57)	(-0.29)	(-0.19)
25th Percentile	0.57	-0.01	-0.40	-0.18	-0.09
Upper C.I. of Random Benchmark	(-0.24)	(-0.24)	(-0.23)	(-0.10)	(-0.05)
50th Percentile	2.18	0.76	0.15	0.11	0.06
Upper C.I. of Random Benchmark	(-0.04)	(-0.03)	(-0.04)	(0.01)	(0.01)
75th Percentile	5.64	2.49	1.12	0.72	0.37
Upper C.I. of Random Benchmark	(0.14)	(0.17)	(0.15)	(0.13)	(0.14)
90th Percentile	16.30	7.14	4.09	2.24	1.54
Upper C.I. of Random Benchmark	(1.48)	(1.30)	(1.41)	(0.67)	(0.59)
			(B) Services		
10th Percentile, Observed	0.06	-1.66	-0.85	-0.49	-0.45
Upper C.I. of Random Benchmark	(-0.43)	(-0.78)	(-0.66)	(-0.36)	(-0.23)
25th Percentile	0.55	-0.11	-0.03	-0.01	-0.05
Upper C.I. of Random Benchmark	(-0.06)	(-0.15)	(-0.17)	(-0.07)	(-0.04)
50th Percentile	2.11	0.31	0.23	0.19	0.05
Upper C.I. of Random Benchmark	(0.02)	(-0.01)	(-0.02)	(0.00)	(0.01)
75th Percentile	7.18	1.68	1.32	1.10	0.62
Upper C.I. of Random Benchmark	(0.50)	(0.18)	(0.14)	(0.14)	(0.16)
90th Percentile	21.56	4.70	5.44	4.15	2.23
Upper C.I. of Random Benchmark	(2.42)	(1.29)	(1.29)	(1.04)	(0.97)

Notes: Table 2 reports industry-level percentiles for $\tilde{p}(d)$ using the discretized densities. Numbers in parentheses are the 5th highest values of the corresponding moments out of 100 counterfactual economies in which plants are randomly assigned to firms without replacement. In order to comply with Census disclosure avoidance requirements, all percentile measures are means of three industries' values around a given percentile.

1.[6] For most industries, the contribution of the within-firm component to spatial agglomeration is small at all spatial scales. For example, for the median manufacturing or service industry, the contribution is around 2% 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 industries. For manufacturing or service 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. Thus there are certainly some industries and spatial scales for which observed total agglomeration has a quantitatively significant within-firm component.

[6] We do not report $p(W|d)$ because the two measures yield very similar results. Rank correlations between $p(W|d)$ and $\tilde{p}(d)$ are above 0.95 for most distance intervals.

Given these findings, we focus on the subset of industries and spatial scales for which the within-firm component has the largest impact on the observed density $f(d)$. Table 3 reports for set of industries for which either $\tilde{p}(20)$ or $\tilde{p}(40)$ is greater than 10%, along with total industry employment, $\tilde{p}(20)$, $f(20)$, $\tilde{p}(40)$, and $f(40)$.[7] Panel (A) reports results for manufacturing industries, which include some canonical examples of highly agglomerated (high $f(d)$) industries, including textile furnishing mills (NAICS 3141, concentrated in the Southeast) and and Motor Vehicle Manufacturing (NAICS 3361 and 3362, concentrated in the Midwest). Panel (B) reports results for service industries, where we also observe some well-known examples of highly agglomerated industries such as several IT industries (NAICS 5112, 5191, and 5415, concentrated in several West Coast cities) and securities brokerages (NAICS 5231, highly concentrated in New York City). Our results indicate that a substantial fraction of the observed agglomeration of these and other industries at short spatial scales can be attributed to the within-firm component.[8] For auto manufacturing and securities brokerages, both well-known for being heavily agglomerated and dominated by a few firms, the within-firm component is particularly significant (about 30% of agglomeration at 0-20 miles).

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 exhibit a significant within-firm component to observed agglomeration, which should be accounted for in studies of agglomeration and coagglomeration that include these industries.

[7] Table A1 reports the set of industries for which $\tilde{p}(20)$ and $\tilde{p}(40)$ are both less than 1%.

[8] Table 3 also included a number of industries (e.g. NAICS 4471, Gas Stations) that have low levels of observed agglomeration (low $f(d)$). In these industries, firms are geographically concentrated even though the industry as a whole is not.

Table 3: List of industries with within-firm contribution > 10% for $\tilde{p}(20)$ or $\tilde{p}(40)$

NAICS	Industry Name	Empl. (k)	$\tilde{p}(20)$	$f(20)$	$\tilde{p}(40)$	$f(40)$
	(A) Manufactures					
3112	Grain and Oilseed Milling	54	13.59	0.007	1.13	0.004
3122	Tobacco Manu.	15	52.21	0.021	4.34	0.031
3131	Fiber, Yarn, and Thread Mills	25	20.57	0.03	6.36	0.029
3141	Textile Furnishings Mills	51	10.10	0.094	12.60	0.017
3162	Footwear Manu.	13	10.81	0.011	8.55	0.004
3212	Veneer, Plywood, and Engineered Wood Prod. Manu.	60	13.41	0.004	13.66	0.003
3253	Pesticide, Fertilizer, and Other Ag. Chem. Manu.	29	22.79	0.008	7.96	0.005
3271	Clay Prod. and Refractory Manu.	34	13.49	0.005	0.51	0.008
3311	Iron and Steel Mills and Ferroalloy Manu.	105	7.79	0.024	25.64	0.022
3313	Alumina and Aluminum Prod.ion and Processing	54	11.70	0.006	-1.51	0.006
3325	Hardware Manu.	27	14.05	0.007	5.47	0.007
3336	Engine, Turbine, and Power Transmission Equip. Manu.	107	15.20	0.008	-0.52	0.004
3361	Motor Vehicle Manu.	150	29.06	0.018	5.16	0.009
3362	Motor Vehicle Body and Trailer Manu.	116	14.96	0.046	10.34	0.01
3364	Aerospace Prod. and Parts Manu.	388	22.26	0.021	2.19	0.008
3365	Railroad Rolling Stock Manu.	29	15.86	0.008	11.43	0.01
3372	Office Furniture (including Fixtures) Manu.	107	14.52	0.011	1.67	0.01
	(B) Services					
4238	Machinery, Equip., and Supplies Wlrs.	698	9.87	0.008	16.96	0.007
4247	Petroleum and Petroleum Prod. Wlrs.	105	18.04	0.01	6.76	0.004
4413	Automotive Parts, Accessories, and Tire Stores	494	12.70	0.004	3.94	0.004
4451	Grocery Stores	2,582	49.99	0.014	50.04	0.012
4461	Health and Personal Care Stores	1,010	D	D	D	D
4471	Gasoline Stations	863	35.69	0.005	29.39	0.005
4529	Other General Merchandise Stores	1,787	D	D	D	D
4533	Used Merchandise Stores	171	13.03	0.006	12.63	0.005
4542	Vending Machine Operators	42	3.61	0.004	14.45	0.005
4831	Deep Sea, Coastal, and Great Lakes Water Tran.	51	10.57	0.067	-1.88	0.021
4832	Inland Water Tran.	20	14.97	0.032	-2.64	0.014
4861	Pipeline Tran. of Crude Oil	10	D	D	D	D
4862	Pipeline Tran. of Natural Gas	33	10.24	0.034	0.22	0.011
4869	Other Pipeline Tran.	9	19.39	0.025	-2.67	0.007
4889	Other Support Activities for Tran.	17	19.17	0.017	0.28	0.007
5112	Software Publishers	397	17.74	0.046	-0.30	0.013
5152	Cable and Other Subscription Programming	51	17.05	0.169	-8.09	0.008
5171	Wired Telecommunications Carriers	756	18.41	0.011	17.69	0.009
5191	Other Information Srvs.	275	10.74	0.035	3.66	0.02
5221	Depository Credit Intermediation	2,009	30.85	0.02	19.90	0.011
5222	Nondepository Credit Intermediation	533	23.25	0.019	11.76	0.009
5231	Sec. and Com. Contracts Intermediation and Brkg.	435	34.84	0.072	-10.20	0.009
5312	Offices of Real Estate Agents and Brokers	247	14.86	0.011	0.09	0.005
5415	Computer Systems Design and Related Srvs.	1,464	10.86	0.031	5.61	0.023
5511	Management of Companies and Enterprises	3,037	33.67	0.022	6.93	0.007
5613	Employment Srvs.	4,971	42.77	0.031	0.63	0.008
5617	Srvs. to Buildings and Dwellings	1,705	21.48	0.012	-2.89	0.005
7131	Amusement Parks and Arcades	165	28.10	0.081	-8.93	0.011
7213	Rooming and Boarding Houses	17	13.28	0.01	10.31	0.003
7225	Restaurants and Other Eating Places	9,025	17.12	0.008	6.92	0.006

Notes: Table reports the list of industries for which $\tilde{p}(d) > 10\%$ for all d, as well as the values of $\tilde{p}(d)$ and $f(d)$. Employment figures are in thousands and are from authors' calculations as well as 2012 County Business Patterns. Entries marked "D" are withheld to comply with Census disclosure avoidance requirements.

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.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg, “The industrial revolution in services,” Technical Report, National Bureau of Economic Research 2019.

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.

Lan, Ting, “The Coagglomeration of Innovation and Production,” Technical Report, mimeo, University of Michigan 2019.

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.

Oberfield, Ezra, Esteban Rossi-Hansberg, Pierre-Daniel Sarte, and Nicholas Trachter, “Plants in space,” Technical Report, National Bureau of Economic Research 2020.

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.

_ and _ , “Evidence on the nature and sources of agglomeration economies,” in “Handbook of regional and urban economics,” Vol. 4, Elsevier, 2004, pp. 2119–2171.

A Appendix

A.1 Least-affected industries

Table A1 provides a list of the industries for which $\tilde{p}(20)$ and $\tilde{p}(40)$ are both less than 1%. In these industries, the within-firm component is a negligible portion of measured industry agglomeration.

A.2 Robustness to Consolidating Nearby Establishments

Our findings in part may come from firms which have multiple establishments within the same 4-digit NAICS industry that are very close. The Census definition of establishments within multi-unit firms requires that establishments be at physically different locations (addresses) to be counted as distinct. To some extent, this precludes the division of production within plants into an arbitrary number of establishments, although there may still be measurement error in the Census' determination of establishment differentiation.

In particular, while there are meaningful economic forces behind the decision to colocate plants very closely, it could be the case that establishments in the same firm at different but extremely close locations do not constitute an organizational structure that is meaningfully distinct from that of a single unified plant. For example, a firm which is space-constrained at an existing establishment may open a new one at an adjoining property with the intention of keeping operations as integrated as possible. The Census may inappropriately categorize these establishments as distinct.

As a purely statistical matter, very close establishments enter into existing agglomeration measures as though they are different companies that colocate. To the extent that they are in effect mis-measured singular plants, this is actually further reason to include these plants in the within-firm adjustment to the industry's total measured agglomeration, since we typically seek to exclude the colocation of employment at single plants from measured industry agglomeration. However, from the perspective of understanding the economic forces pushing firms to colocate truly distinct production processes nearby, it may be of interest to investigate how our results differ when we combine establishments that are very close.

Tables A2 and A3 replicate our main results while consolidating firms' establishments in the same zipcode and 4-digit NAICS. Necessarily, this re-categorization reduces $\tilde{p}(20)$ and $f^w(20)/f^a(20)$, as a significant portion of plants which both share a firm and are within 20 miles of each other are also within the same zipcode, which are generally several miles wide. Removing these inter-establishment links reduces both $\tilde{p}(20)$ and $f^w(20)/f^a(20)$ by roughly one-third on average. Still, these numbers remain well above the random benchmark and a substantial proportion of agglomeration within 20 miles is within-firm (as well as across zipcodes) in a smaller subset of industries. Because this exercise reduces the total number of unique cross-plant employment pairs, there's also a very slight increase for measures at higher distances.

We caution that these results are an extreme check to our main results. For sure, many firms operate multiple, truly separate operations within the same zipcode. Even excluding these, within-firm agglomeration appears to be a meaningful contribution to some industries' measured agglomeration. Furthermore, to the extent that our main results include inappropriately divided unique plants, they constitute the correct adjustment for the extent to which measured agglomeration is within-firm.

Table A1: List of industries with within-firm contribution $< 1\%$ for $\tilde{p}(20)$ and $\tilde{p}(40)$

NAICS	Industry Name	Empl. (k)	$\tilde{p}(20)$	$f_d(20)$	$\tilde{p}(40)$	$f_d(40)$
	(A) Manufactures					
3369	Other Transportation Equipment Manufacturing	35	-1.02	0.007	0.17	0.004
3118	Bakeries and Tortilla Manufacturing	285	0.07	0.01	0.20	0.007
3121	Beverage Manufacturing	136	-0.31	0.01	-0.39	0.008
3152	Apparel Knitting Mills	95	0.04	0.161	0.03	0.031
3161	Leather and Hide Tanning and Finishing	4	-0.95	0.024	-0.95	0.006
3222	Converted Paper Product Manufacturing	247	0.68	0.007	-0.21	0.006
3231	Printing and Related Support Activities	472	0.66	0.008	0.38	0.007
3259	Other Chemical Product and Preparation Manufacturing	84	0.51	0.006	0.05	0.006
3322	Cutlery and Handtool Manufacturing	38	0.21	0.008	-0.37	0.009
3327	Machine Shops; Turned Product; and Screw, Nut, and Bolt Manufacturing	385	0.28	0.008	0.24	0.008
3328	Coating, Engraving, Heat Treating, and Allied Activities	124	0.48	0.013	0.25	0.01
3329	Other Fabricated Metal Product Manufacturing	260	0.95	0.006	0.66	0.006
3335	Metalworking Machinery Manufacturing	145	0.27	0.015	0.05	0.012
3341	Computer and Peripheral Equipment Manufacturing	67	-2.76	0.015	-2.73	0.005
3342	Communications Equipment Manufacturing	99	0.65	0.014	-0.77	0.015
3346	Manufacturing and Reproducing Magnetic and Optical Media	15	-1.58	0.009	-1.58	0.006
3366	Ship and Boat Building	136	-6.03	0.022	-5.32	0.011
3379	Other Furniture Related Product Manufacturing	32	-1.80	0.009	-1.66	0.008
3391	Medical Equipment and Supplies Manufacturing	300	0.47	0.01	0.21	0.009
	(S) Services					
4231	Motor Vehicle and Motor Vehicle Parts and Supplies Merchant Wholesalers	367	0.55	0.008	0.26	0.007
4232	Furniture and Home Furnishing Merchant Wholesalers	142	0.16	0.02	0.04	0.014
4233	Lumber and Other Construction Materials Merchant Wholesalers	189	0.16	0.007	0.21	0.006
4235	Metal and Mineral (except Petroleum) Merchant Wholesalers	154	0.26	0.013	-0.04	0.009
4237	Hardware, and Plumbing and Heating Equipment and Supplies Merchant Wholesalers	225	0.42	0.008	0.11	0.006
4239	Miscellaneous Durable Goods Merchant Wholesalers	344	0.08	0.015	0.03	0.011
4241	Paper and Paper Product Merchant Wholesalers	149	0.17	0.012	-0.04	0.008
4243	Apparel, Piece Goods, and Notions Merchant Wholesalers	192	0.04	0.093	0.01	0.031
4244	Grocery and Related Product Merchant Wholesalers	756	0.54	0.014	-0.23	0.008
4246	Chemical and Allied Products Merchant Wholesalers	151	0.44	0.011	-0.21	0.01
4249	Miscellaneous Nondurable Goods Merchant Wholesalers	336	0.36	0.008	0.31	0.007
4412	Other Motor Vehicle Dealers	125	0.53	0.003	0.50	0.004
4452	Specialty Food Stores	155	0.54	0.013	0.41	0.01
4483	Jewelry, Luggage, and Leather Goods Stores	130	-0.48	0.008	-0.44	0.008
4543	Direct Selling Establishments	188	-0.02	0.004	0.25	0.007
4854	School and Employee Bus Transportation	211	0.22	0.015	-2.23	0.02
4859	Other Transit and Ground Passenger Transportation	96	0.30	0.024	0.30	0.018
4879	Scenic and Sightseeing Transportation, Other	3	-1.47	0.089	-1.53	0.01
4883	Support Activities for Water Transportation	110	-0.93	0.025	-2.08	0.007
4884	Support Activities for Road Transportation	91	0.41	0.006	0.29	0.007
4921	Couriers and Express Delivery Services	498,000	D	D	0	
4931	Warehousing and Storage	692,000	0.92	0.006	-0.33	0.006
5242	Agencies, Brokerages, and Other Insurance Related Activities	855,000	0.90	0.008	0.63	0.006
5251	Insurance and Employee Benefit Funds	125,000	0.11	0.023	0.00	0.01
5311	Lessors of Real Estate	537,000	0.77	0.018	0.45	0.008
5321	Automotive Equipment Rental and Leasing	145,000	-1.40	0.009	-0.50	0.005
5324	Commercial and Industrial Machinery and Equipment Rental and Leasing	156,000	0.88	0.009	-0.03	0.007
5411	Legal Services	1,152,000	0.13	0.019	0.42	0.009
5412	Accounting, Tax Preparation, Bookkeeping, and Payroll Services	1,297,000	0.74	0.006	0.63	0.006
5611	Office Administrative Services	490,000	0.21	0.012	0.05	0.008
5616	Investigation and Security Services	845,000	-0.30	0.017	0.22	0.011
5619	Other Support Services	275,000	0.25	0.011	-0.52	0.008
5629	Remediation and Other Waste Management Services	130,000	0.15	0.007	0.15	0.006
7111	Performing Arts Companies	117,000	0.81	0.035	-0.02	0.008
7112	Spectator Sports	130,000	0.80	0.01	0.09	0.006
7114	Agents and Managers for Artists, Athletes, Entertainers, and Other Public Figures	18,000	-0.20	0.154	-0.35	0.02
7115	Independent Artists, Writers, and Performers	48,000	0.00	0.049	0.00	0.014
7223	Special Food Services	674,000	0.34	0.027	-0.01	0.022
7224	Drinking Places (Alcoholic Beverages)	350,000	0.15	0.008	0.01	0.004
8114	Personal and Household Goods Repair and Maintenance	68,000	0.21	0.007	-0.30	0.007
8131	Religious Organizations	1,686,000	0.01	0.005	0.00	0.005
8133	Social Advocacy Organizations	180,000	0.39	0.021	0.06	0.007
8139	Business, Professional, Labor, Political, and Similar Organizations	498,000	0.05	0.02	0.00	0.007

Note: Table reports the list of industries for which $\tilde{p}(d) < 1\%$ for $d = 20$ and $d = 40$. Employment figures are in thousands and are from authors' calculations as well as 2012 County Business Patterns. Entries marked "D" are withheld to comply with Census disclosure avoidance requirements.

Table A2: Ratio of within to across firm densities, consolidated by zipcode, by percentile

Distance	0-20 miles	20-40 miles	40-60 miles	60-160 miles	160-300 miles
			(A) Manufactures		
10th Percentile, Observed	0.47	0.28	0.03	0.26	0.51
Upper C.I. of Random Benchmark	(0.00)	(0.01)	(0.01)	(0.37)	(0.51)
25th Percentile, Observed	1.23	0.98	0.51	0.77	0.82
Upper C.I. of Random Benchmark	(0.18)	(0.25)	(0.26)	(0.70)	(0.81)
50th Percentile, Observed	3.23	1.82	1.44	1.38	1.15
Upper C.I. of Random Benchmark	(0.65)	(0.73)	(0.76)	(1.00)	(1.07)
75th Percentile, Observed	7.62	4.55	2.96	1.89	1.45
Upper C.I. of Random Benchmark	(1.38)	(1.49)	(1.46)	(1.38)	(1.37)
90th Percentile, Observed	14.04	7.15	5.36	3.22	1.87
Upper C.I. of Random Benchmark	(2.78)	(3.08)	(3.31)	(2.17)	(2.01)
			(B) Services		
10th Percentile, Observed	0.85	0.16	0.25	0.38	0.34
Upper C.I. of Random Benchmark	(0.26)	(0.07)	(0.04)	(0.36)	(0.55)
25th Percentile, Observed	1.27	0.85	0.91	1.02	0.89
Upper C.I. of Random Benchmark	(0.70)	(0.51)	(0.52)	(0.78)	(0.86)
50th Percentile, Observed	2.59	1.62	1.54	1.33	1.13
Upper C.I. of Random Benchmark	(1.01)	(0.90)	(0.90)	(0.99)	(1.02)
75th Percentile, Observed	7.34	4.55	3.03	2.57	1.62
Upper C.I. of Random Benchmark	(1.51)	(1.24)	(1.20)	(1.19)	(1.20)
90th Percentile, Observed	18.99	11.28	7.32	6.19	2.61
Upper C.I. of Random Benchmark	(2.70)	(1.90)	(1.82)	(1.63)	(1.65)

Note: Table replicates Table 1, but treats all establishments in the same firm and zipcode as a single establishment.

Table A3: Counterfactual contribution of within-firm component consolidated by zip-code, by percentile

Distance	0-20 miles	20-40 miles	40-60 miles	60-160 miles	160-300 miles
			(A) Manufactures		
10th Percentile, Observed	-1.12	-0.44	-1.53	-1.15	-0.36
Upper C.I. of Random Benchmark	(-0.58)	(-0.59)	(-0.54)	(-0.29)	(-0.19)
25th Percentile, Observed	0.15	-0.01	-0.39	-0.17	-0.09
Upper C.I. of Random Benchmark	(-0.26)	(-0.24)	(-0.22)	(-0.11)	(-0.05)
50th Percentile, Observed	1.10	0.79	0.15	0.12	0.07
Upper C.I. of Random Benchmark	(-0.05)	(-0.04)	(-0.04)	(0.00)	(0.01)
75th Percentile, Observed	3.08	2.50	1.15	0.72	0.38
Upper C.I. of Random Benchmark	(0.09)	(0.13)	(0.17)	(0.15)	(0.15)
90th Percentile, Observed	13.14	7.17	4.10	2.25	1.61
Upper C.I. of Random Benchmark	(1.17)	(1.06)	(1.23)	(0.65)	(0.57)
			(B) Services		
10th Percentile, Observed	-0.13	-1.39	-0.73	-0.40	-0.29
Upper C.I. of Random Benchmark	(-0.52)	(-0.72)	(-0.61)	(-0.33)	(-0.19)
25th Percentile, Observed	0.12	-0.05	-0.01	0.00	-0.03
Upper C.I. of Random Benchmark	(-0.09)	(-0.15)	(-0.15)	(-0.07)	(-0.03)
50th Percentile, Observed	0.92	0.34	0.24	0.21	0.06
Upper C.I. of Random Benchmark	(0.00)	(-0.02)	(-0.02)	(0.00)	(0.00)
75th Percentile, Observed	3.38	1.70	1.39	1.17	0.67
Upper C.I. of Random Benchmark	(0.33)	(0.19)	(0.16)	(0.14)	(0.13)
90th Percentile, Observed	13.21	4.72	5.50	4.21	2.35
Upper C.I. of Random Benchmark	(1.92)	(1.15)	(1.15)	(0.96)	(0.91)

Note: Table replicates Table 2, but treats all establishments in the same firm and zipcode as a single establishment.