Productivity, Place, and Plants: Revisiting the Measurement*

Benjamin Schoefer
UC Berkeley

Oren Ziv
Michigan State University

Preliminary. Update of empirical results pending due to ongoing closure of Census RDCs since March 2020.
Currently, some entries are missing, and some will be replaced.

July 25, 2020

Abstract

We show that most of the large measured geographic dispersion in productivity across US cities is spurious and reflects granularity: idiosyncratic plant-level productivity variation combined with finite plant counts. As a result, economies with randomly reallocated plants exhibit nearly as high a variance as the empirical economy. Our split-sample strategy, which nonparametrically strips out the granularity bias, reduces the raw variance of place effects by two-thirds. For new plants, seven eights of the dispersion of place effects reflects noise. These patterns extend to 15 European countries.

*We thank Luke Watson for excellent research assistance. We thank Gabriel Chodorow-Reich (discussant), Jonathan Dingel, and Cecile Gaubert for useful discussions. We thank partcipants of the NBER Summer Institute Urban Economics Meeting for feedback. We thank the Ewing Marion Kauffman Foundation for financial support. The work draws on confidential data from the US Census Bureau, which requires the following disclaimer: "Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the US Census Bureau. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1890. All results have been reviewed to ensure that no confidential information is disclosed."

1 Introduction

The sources and consequences of spatial differences in productivity are a frequent focus in urban economics. They contribute to regional dispersion in local labor market outcomes, for example raising wages and employment (e.g., Caliendo, Parro, Rossi-Hansberg, and Sarte, 2018; Hornbeck and Moretti, 2018), and are considered a driver of plant location choice and a testable implication of core of urban economics models (e.g., Fajgelbaum, Morales, Suárez Serrato, and Zidar, 2019; Gaubert, 2018; Davis and Dingel, 2020). Furthermore, these differences motivate studies of spatial misallocation and the design of spatial policies (Moretti, 2012; Fajgelbaum and Gaubert, 2018; Boeri, Ichino, Moretti, and Posch, 2019; Fajgelbaum, Morales, Suárez Serrato, and Zidar, 2019; Hsieh and Moretti, 2019).

We revisit the measurement of the spatial dispersion in productivity, using US Census data covering the universe of manufacturing plants. Our headline number is the variance of average log productivity (TFP) across cities (metropolitan statistical areas, MSAs) demeaned by industry. This variance is high: 0.024, implying large spatial dispersion. For example, manufacturing plants in the top 5% most productive Metropolitan Statistical Areas (MSAs) appear roughly two thirds (0.5 log points) more productive than plants in the bottom 5%. We also construct finer industry-specific location measures, which exhibit double the variance.

Rather than reflecting dispersion in true place effects, we show that these high raw variances spuriously arise from tremendous idiosyncratic plant-level heterogeneity averaged over finite plant counts. True place effects denote a plant's *expected* (log) productivity in a given location. This statistical definition of place effects is agnostic to their sources, encompassing potential causal effects, sorting, or spatially correlated mismeasurement. Yet, the observed location-specific *averages* of plants' heterogeneous productivities features considerable sampling variability, which enters the raw variances of place averages as a term we call granularity bias.

We assess the granularity bias in two complementary strategies. First, we provide a nonparametric permutation test to assess whether the empirical place effects do at all differ statistically from a random allocation of plants across space, an exercise in the spirit of the dartboard approach in Ellison and Glaeser (1997). We ask: what is the probability that the observed level of spatial dispersion emerged if plant-level productivity were independent of place? We simulate 1,000 US economies that randomly reassign the empirical plants across MSAs. Their variances in place effects make up our nonparametric sampling distribution. The average variance is 0.016, which is high, but one-third below the empirical number (0.024), which however is in the 99th percentile the distribution of the randomized economies. For industry-specific location effects, the empirical variance of 0.048 is close to the permutation-based variance of 0.043. But it is in the 96.7th pecentile of the distribution of permuted variances. Overall, our exercise clarifies that despite the capacity of granularity to spuriously generate substantial spatial variation even in randomized economies, the real US economy exhibit a statistically significant degree of true place and place-industry effects.

Second, we constructively provide a bias-corrected estimate of the empirical variance using a split-sample approach. In each location, we randomly split plants into two groups, and estimate separate place effects. The covariance between the twin cities' place effects is an unbiased estimator of the variance of true place effects, which are common to both split samples. This cleansed variance shrinks substantially, from 0.024 to 0.008, and from 0.048 to 0.013 for the industry-specific variant. In robustness checks, we find similar results for labor productivity (value added per worker) rather than TFP, using different industry definitions, doubling the minimum cell-level count from 5 to 10, studying only older plants, and not weighting plants by employment.

Hence, granularity bias accounts for two thirds of the raw variance of place effects. The remaining dispersion in true place effects arises from at least three sources: (i) causal effects of place on productivity, (ii) sorting among ex-ante differentiated plants (Gaubert, 2018), and (iii) remaining measurement error common to a location. Our bias-corrected estimates lower the upper bound for these sources. It may also apply to those calculations of the degree of spatial misallocation that draw on measured productivity dispersion.

We further investigate the dispersion in place effects for new plants. Place effects among the population of plants are dominated by older plants, which may reflect legacy technologies. Moreover, certain places may be specifically productive for new plants. TFP in new projects may also be more relevant for entrepreneurship, plant location choices, and marginal investment and hiring decisions more generally. We find that place effects measured from new plants (at most five years old) are markedly more dispersed in the raw data, with a variance of 6.0%, and even higher at 13% when we permit industry-specific place effects. Cleaning these raw variances using our split-sample approach, the variance of true place effects of new plants drops to 0.8% for place effects, and 1.4% respectively. Hence, bias from granularity even further overstates the dispersion of place effects for new plants. Moreover, we show that the baseline raw place effects constructed using all plants are even less informative for new plants, though they do have some predictive power.

We close with a second application of our methods to a series of 15 European countries and their within-country regional dispersion, drawing on Bureau van Dijk firm-level data in the manufacturing sector. As in the US, in most of these countries, the spatial variation in productivity is tremendously reduced once accounting for granularity bias.

2 Statistical Basics: Place Effects Under Granularity

In this section, we clarify the pitfalls of estimating dispersion of place effects and present our strategies to overcome these measurement challenges.

2.1 Setting

The economy is characterized by a set L of N_L locations indexed by $l \in L$. Each location has N^{P_l} plants indexed by $p \in P_l$, where P_l denotes the population of location-l plants; we will also consider subsets $S_l \subseteq P_l$ of size N^{S_l}.

Plant p in l has log productivity $a_{pl} = \ln A_{pl}$. We leave open whether a_{pl} captures true TFP or also capture mismeasurement. Plants are heterogeneous in productivity in ways that potentially depend on their location l, i.e. $a \sim F^a_l(a)$. In addition, a plant is characterized by size e_p; the exposition below starts with unweighted (or equally sized) plants within locations, and no weighting across locations. We discuss the extension to weighted averages and heterogeneity in size in Section 2.5.

True Place Effects We define true place effects as *expected values* of plant-level productivity in location l:

$$\tau_l = \mathsf{E}\left[a_{pl} \middle| l\right] = \int a dF_l(a). \tag{1}$$

This *statistical* definition of the place effects is agnostic to and maps into a variety of specific economic mechanisms. It may, for instance, capture causal effects of place on productivity, including from agglomeration effects. Alternatively, it may capture systematic sorting or collocation of plants into places by productivity. It may also capture location-specific mismeasurement of productivity (e.g., in the production functions, input and output prices, or input quantities or qualities). The place effect draws on a_{pl} i.e. productivities of existing plants after plant location choices and entry and exit dynamics, so it may also capture productivity-relevant spatial differences therein. Of course, our formulation of place effects as expected values would not sufficiently characterize any given specific model or mechanism; instead, $F^a_l(a)$ may differ across places $l \in L$ in various moments due to the aforementioned factors.

Our goal is to characterize the dispersion of true place effect τ_l across locations $l \in L$.

Idiosyncratic Plant-Level Productivity Using our definition of place effects in Equation (1), we can rewrite a plant's log productivity a_{pl} as the sum of the place effect τ_l and an idiosyncratic component u_{pl}:

$$a_{pl} = \tau_l + u_{pl}. \tag{2}$$

That is, plants' idiosyncratic residuals within a location l are simply deviations around that expected value τ_l drawn from a potentially location-specific distribution $u \sim F_l^u(u)$. Hence, their expected value is zero: $\mathsf{E}\left[u_{pl}|l\right] = 0$. Moreover, $F_l^a(a) = F_l^u(a - \tau_l)$. We take no stance on the economic origin of these deviations. Like plant productivity a_{pl}, idiosyncratic deviation u_{pl} may capture true productivity deviations of heterogeneous plants or plant-specific measurement error (e.g., Bils, Klenow, and Ruane, 2020).

Average-Based Place Effects As $\mathsf{E}\left[u_{pl}|p \in S_l\right] = 0$ for any random sample S_l, the average productivity for those plants $p \in S_l$ is an unbiased and consistent estimator of τ_l:

$$\widehat{\tau}_l^{S_l} = \frac{1}{N^{S_l}} \sum_{p \in S_l} a_{pl}. \tag{3}$$

Even with population data of finite plants $S_l = P_l$, as with our census data, the average defined in Equation 3 generally differs from our object of interest, the expectation $\mathsf{E}(a_{pl}|l)$ defined in Equation 1. That expectation is taken over the latent data-generating process $F_l^a(a)$ of plant productivities, from which the real economy and the census data draw a finite set of plants.

2.2 Pitfalls of Estimating Dispersion in Place Effects

While each place effect $\widehat{\tau}_l$ is estimated without bias, dispersion measures based on averages $\widehat{\tau}_l$ are upward-biased estimates of the dispersion of true place effects τ_l, for reasons we label "granularity." Below we formalize this at the example of our leading dispersion statistic, namely the variance.

Variances of Place Averages Formally, the pitfalls of estimating place effects on the basis of location averages become clear when considering the variance of the place effect estimators i.e. that of averages $\widehat{\tau}_l^{S_l}$:

$$\overbrace{\mathsf{Var}(\widehat{\tau}_l^{S_l})}^{\substack{\text{Var of Est. Place Effects}\\ \text{(Location Averages)}}} = \mathsf{Var}\left(\frac{1}{N^{S_l}}\sum_{p\in S_l}[\tau_l + u_{pl}]\right) = \mathsf{Var}\left(\tau_l + \frac{1}{N^{S_l}}\sum_{p\in S_l} u_{pl}\right) \tag{4}$$

$$= \underbrace{\mathsf{Var}(\tau_l)}_{\text{Var of True Place Effects}} + \underbrace{\frac{1}{L}\sum_{l\in L}\frac{\sigma_l(u)^2}{N^{S_l}}}_{\substack{>0 \text{ if } N<\infty\wedge\sigma(u)>0\\ \text{Bias from Granularity:}\\ \text{Var of Sample Means}}} + \underbrace{2\,\mathsf{Cov}\left(\tau_l, \frac{1}{N^{S_l}}\sum_{p\in S_l} u_{pl}\right)}_{\substack{=0\\ \text{Orthogonal by Construction}}}, \tag{5}$$

where $\sigma_l(u)^2$ is the variance of plant-specific deviations u_{pl} from place effect τ_l in location l. The third term is zero because the idiosyncratic deviations from the expected value are orthogonal to the expected value. The remaining terms are the true variance plus our bias term, generating excess variance due to granularity.

Granularity While the variance of place averages $\widehat{\tau}_l^{S_l}$ is a consistent estimator of the variance of true place effects τ_l, with finite samples of plants within locations it is *biased upward* by the second term: the weighted average of within-location variances divided by location count of plants. Formally, this bias term in Equation 5 arises from "granularity:" small cell counts of plants N^{S_l} combined with large idiosyncratic variance $\sigma_l(u)^2$ within a cell l and is present even when $S_l = P_l$.[1] Intuitively, these factors generate realized deviations of sample averages from τ_l, increasing the overall observed variance of place averages.

To empirically gauge N_l, we plot the CDF of cell sizes (plant counts) from a public-use data on manufacturing plants in the US County Business Patterns in Figure A.1. Panel (a) plots the CDF of plant counts for cells defined as MSAs industry (pooling all industries); Panel (b) does so for MSAs 86 4-digit NAICS cells. While small cells are an obvious issue in measuring variance across industry-locations, for which 60% have fewer than 5 plants, there are more plant observations at the MSA level when pooling all industries. Importantly, Equation 5 clarifies that even for larger N_l, granularity bias can be large if plants exhibit large idiosyncratic variance in at least some locations, so that even MSA effects pooling all industries can exhibit considerable bias. We will separately study location and industry-location cell definitions.

[1]The setting is "granular" in the sense that a given plant need not wash out in the average. Equation 5 permits heteroskedasticity in the plant distribution across locations $F_l(u)$, mirrored in the l-index of $\sigma_l(u)$. Gaubert and Itskhoki (forthcoming) explore granular comparative advantage in international trade. Dingel and Tintelnot (2020) study granularity in a spatial general equilibrium setting with commuting.

2.3 Permutation Test: Pure Granularity and No Place Effects

We formulate a clean, extreme benchmark for the distribution of place effects: that all locations $l \in L$ have the same data-generating process for plant productivity $F_l^a(a) = F^a(a)\ l \in L$ and hence have the same expected value $\tau_l = \tau \forall l \in L$, so that the variance of true place effects is zero. Dispersion in measured place averages arises solely as an artifact of grouping heterogeneous plants, i.e. from granularity bias.

A potential statistical test of this hypothesis, however requiring parametric assumptions, is an F-test of joint significance of all place averages (for example, estimated as fixed effects in a regression) being zero.

Instead, we implement a nonparametric (or randomization), exact test of this hypothesis in the spirit of permutation tests. We construct the sampling distribution of our test statistic of interest under the following procedure: plants are randomly distributed across space into MSAs. Specifically, we preserve the count of plants by each industry-location. Under this procedure, the rank of a given empirical dispersion static in the CDF of those of the random economies gives the nonparametric p-value corresponding to that null hypothesis.

Broadly, by referencing a random-location benchmark, our test of productivity place effects is in the spirit of Ellison and Glaeser (1997), who study whether the observed geographical concentration is statistically different from randomly located plants.

2.4 Bias Correction of Variance: Split Samples

To provide a constructive measure of the dispersion in true place effects, we implement a split-sample procedure to remove the bias, essentially using one half of the plants within a given MSA as an instrument for the other half. Formally, we split the plants into two random and equally sized subsamples $s \in \{A, B\}$ in each location l, essentially generating a new location label l, A and l, B. We then estimate place effects $\widehat{\tau_l}^A, \widehat{\tau_l}^B$ for all cells (l, s). We then calculate the *covariance* of the two separate sets of fixed effects across locations l between half-samples A and B:

$$\begin{aligned}
\mathsf{Cov}\left(\widehat{\tau}_l^A, \widehat{\tau}_l^B, \right) &= \mathsf{Cov}\left(\tau_l + \bar{u}_l^{A_l}, \tau_l + \bar{u}_l^{B_l}\right) &(6)\\
&= \mathsf{Var}\left(\tau_l\right) + \underbrace{\mathsf{Cov}\left(\tau_l, \bar{u}_l^{A_l}\right)}_{=0} + \underbrace{\mathsf{Cov}\left(\tau_l, \bar{u}_l^{B_l}\right)}_{=0} + \underbrace{\mathsf{Cov}\left(\bar{u}_l^{A_l}, \bar{u}_l^{B_l}\right)}_{=0}, &(7)
\end{aligned}$$

where we have introduced the notation of $\bar{u}_l^{S_l} = \frac{1}{N^{S_l}} \sum_{p \in S_l} u_{pl}$ as the sample average of deviations u_{pl} in a sample S_l of location l. Because subsamples are drawn randomly in our procedure, the second and third terms are zero. To ensure the fourth term is zero, we introduce an assumption, that errors $u_{pl}, u_{p'}$ are independent within locations l.

Hence, the covariance of averages of randomly chosen subsamples is an unbiased an estimator of the variance of the true place effects, eliminating granularity bias.[2]

2.5 Weighting

First, consistent with our empirical implementation, the variance does not weight across locations, so the bias term is the unweighted average of all location's l-specific bias.

Second, the expressions so far have presented the case without weighting plants within a location; equivalently with homogeneous sizes. In practice, we follow the literature by weighting plants by a plant employment e_{pl} when constructing cell-level averages. The true place effect then takes the expectation of productivity over the joint distribution of plant productivity and size $F^l(u, e)$. The main implication is that the bias from granularity now also encompasses the potential dominance of large plants. The covariance remains the unbiased estimator of the variance of true place effects. We additionally (in Appendix Table A.1) present specifications in which place effects are constructed without weighting by plant size. The unweighted results yield smaller raw variances; yet the bias correction strips out a similar share of the variances, such that the bias-corrected variances from the weighted specification remain higher by around 50-80%.

3 Data and Construction of Place Effects

Plant-Level Data: US Census of Manufactures Our primary data set is plant-level data on production and plant characteristics for the universe of manufacturing plants responding to five bidecadal Census waves of the US Census of Manufactures (CMF) from 1992-2012. We observe plants' MSA, which we take as the basic unit of location, as well as their 6-digit NAICS code (following the time consistent definition of Fort and Klimek, 2018). For our primary specifications, we consider industry-location cells consisting of 4-digit industry groups; we probe robustness to using the 6-digit classification and report results in Appendix Table A.1. Hence, we observe plants in location-industry cells, which potentially divide our sample into 361 MSAs and 86 4-digit industries, before applying our minimum cell size restrictions below.

[2]This adjustment has been used in the context of group-level wage differences (namely, AKM firm wage fixed effects Gerard, Lagos, Severnini, and Card, 2018; Drenik, Jäger, Plotkin, and Schoefer, 2020), or to estimate peer effects in personnel economics (Silver, 2020). We are not aware of applications to the measurement of group-level averages of productivity. Bils, Klenow, and Ruane (2020) use a within-firm IV strategy to adjust for measurement error in within-plant productivity.

For each plant, we observe revenue, and inputs (employment (separately of production and non-production workers), materials, capital expenditures and energy expenditures. Value added is revenue minus non-labor inputs. We also construct plant age, as the difference between survey year and the first time the plant enters the the Longitudinal Business Database (LBD), which tracks plants on an annual basis since 1977.

Cell Size We do not apply plant-level sample restriction except that we require TFP and value added per worker to be computable; instead, we select our sample on the cell level, keep industry-location cells with at least 5 plants. This leaves us with roughly 1,500 MSA-year cells and 15,000 industry-MSA-year cells. As a robustness check, we increase these cutoffs to 10 plants, with results reported in Appendix Table A.1.[3][4]

Plant-Level TFP Assuming a standard Cobb-Douglas production function with constant returns, a plant p's TFP(r) is the residual of inputs subtracted from output, with factor shares of its industry $i(p)$.[5] We follow and use the plant-level TFPr construction of Foster, Haltiwanger, and Syverson (2008) on the basis of industry-year input and output deflators. We winsorize the final plant-level TFP measures at 1% and 99%.

Location-Industry Effects We start by defining industry-specific location effects, denoting these by $\tau_{l,i}$, demeaned within each Census year and industry. This measure captures the percent (log) premium of plants in the location-industry compared to (demeaned by) their industry peers nationally.

In practice, we generate these averages in a fixed-effects regression with location-industry fixed effects, weighted by plant employment. In a second step, we demean these industry-location fixed effects nationally (without weighting).[6]

[3]Specifically, we keep industry-location cells with 2 plants five years or younger and at least 3 plants over 5 years of age, the minimum number of plants needed to complete our analysis of old and young plants (so that our robustness check of 10 plants requires 6 old and 4 young plants) In a future version, we will probe robustness to keeping cells with at least 5 plants of any age for the main analyses.

[4]This eliminates roughly 90% of our potential cells. Because TFP is calculable for roughly half of plants in the the Census of Manufactures, when compared to CBP cell counts, we would expect this number to correspond to the cumulative density at 10 in Appendix Figure A.1, recalling that the empty 37% of cells are not represented there.

[5]Industries $i \in I$ are defined as groups of plants p that share a common Cobb-Douglas production function, such that plan p produces output $Y_p = A_p \prod_\iota {q_p^\iota}^{c^\iota_{i(p)}}$, so that TFP is $A_p = \frac{Y_p}{\prod_\iota {q_p^\iota}^{c^\iota_{i(p)}}}$, where q_p^ι is plant p's input quantity of type ι, and c_i^ι the industry-specific factor share. Here, the units are real quantities, so A_p real total factor productivity (TFPq). In practice, we will measure TFPr, using a common industry price deflator, such that we will absorb price differences between plants within an industry into our TFP measure.

[6]In principle, the construction of these cell-level deviations from the industry average is subject to the measurement issues related to granularity too, which we sidestep.

Location Effects As our most comprehensive measure, we collapse (average) the industry-place effects $\tau_{i,l}$ into one place effect ξ_l across all industries. To make this notion comparable with the industry-location effects, we weight by the industry's employment share within an MSA among the cells surviving the restrictions.[7] In a final step, we center these place effects around zero nationally by demeaning them.

In practice, we estimate these place effects in a fixed-effects regression with location fixed effects and industry fixed effects, where we weight by plant employment.

Weighting As a robustness check, Appendix Table A.1 replicates our results with equally weighted plants.

One Vs. Multiple Census Waves Our preferred specification would take one census cross section. Our currently disclosed results repeat the exercise separately for each Census year (i.e., all fixed effects are year-specific) – and pool all such years (such that we effectively have five US economies (one per Census year), which we pool). [We will reduce the Census to the most recent, i.e. 2012 wave after the Covid closures of the Census RDCs.]

Alternative Productivity Measure: Value Added per Worker We complement our primary productivity measure, TFP, with a less demanding measure in the form of value added per worker. We report these results in Appendix Table A.1. This measure does not require specifying a production function and its parameters, and measuring each input for each firm. As a benchmark, with Cobb-Douglas production, the marginal product of labor would correspond to the labor share in production (counting all production inputs) times output per worker, or the labor share in a labor and capital income only times value added per worker. We take logs and include industry effects, which net out the labor shares if constant in the industry.

Additional Data for Extension to European Countries: Bureau van Dijk Firm-Level Data We complement our analysis of US plant-level Census data with firm-level data from Bureau van Dijk covering 15 European countries. On a country-by-country basis, we replicate our construction of place and industry-place effects using NUTS-2 within-country regional divisions. Each NUTS-2 region contains between 800,000 and 3 million inhabitants. We construct TFP measures for the manufacturing sector, at the 2-digit NACE

[7]Employment weights are commonly used in the literature to weight within and across industries (e.g., Hornbeck and Moretti, 2018). There is a slight inconsistency between location-industry and location effects in the way the national industry effects are constructed (weighted by plant employment in the location effects, but weighted by local employment share in the location-industry effects.

industry level, following the approach detailed in Jäger, Schoefer, and Heining (2019).[8] We winsorize the resulting TFP measures at 1% and 99%.

These data have several drawbacks for our purposes. Unlike a census or tax data, they have imperfect coverage, varying data quality specifically regarding value added (see, e.g., Gopinath et al., 2017). To maximize coverage and mimic a census, we keep each firm's most recent observation, implying that most observations come from the late 2010s. The capital stock measure is based on reported book value of assets. We do not apply industry-specific input price indices (industry fixed effects absorb national output price indices). Unlike our US plant data, BvD data is firm-level, and multi-plant firms with potentially many industries and locations are assigned to a single industry and headquarter location. As with our US data, we keep location-industry cells with at least 5 firms. Appendix Table A.2 lists number of regions, cells, and average number of firms per cell by country.

4 The Dispersion in Productivity Across US Locations

We first measure the naive geographic dispersion in productivity. We then implement our permutation test to assess the null that hypothesis the the empirical variance is entirely spurious. Finally, we implement our proposed adjustment to cleanse the naive dispersion of granularity bias.

4.1 Raw Dispersion

Figure 1 plots the distribution of place effects for locations and location-industries.

Location Effects ξ_l Figure 1 Panel (a) plots the distribution of location-specific effects ξ_l.[9] The focus of this section is the black solid line, the raw distribution of place effects. The location effects ξ_l trace out a bell-shaped distribution.[10] As printed into Figure 1, the variance is 0.024. This statistic is our headline number reported in the introduction in Section 1. That is, plants in MSAs with location effects that are one standard deviation above the mean have on average around 15% higher productivity than plants in the same

[8]That is, we construct the industry-country-specific labor share by dividing the the sum of payroll at the industry level of all sample firms by the corresponding sum of value added for firms with nonmissing observations on both variables. We then construct firm-specific TFP by assuming a Cobb-Douglas production function, and subtracting from log value added labor-share weighted employment and one-minus-labor-share-weighted log capital. We use fixed tangible assets as the capital stock measure.

[9]The graphs present kernel densities, as the Census disclosure process does not permit extracting raw data histograms.

[10]The effects need not be centered around zero because we do not weight across MSAs by employment, while the demeaning within industry at the plant-level regression is employment-weighted.

industry nationally. The productivity difference between the 5th and 95th percentiles is estimated to be around 65% (0.5 log points).

Location-Industry Effects $\tau_{i,l}$ Panel (b) repeats this analysis for the location-industry effects $\tau_{i,l}$, which permit place effects to vary by industry. We weight these values by the industries' employment share in the location, for comparability with Panel (a). As a benchmark, if all industries in a location deviated by the same percent from their national industry benchmark, then the location and location-industry effects would exhibit the same dispersion.

Instead, we find a considerably more dispersed distribution of the industry-location effects, with a variance of 4.8%. [Panel (b) is currently based on rescaled data with pending year-industry demeaning, following the method in Footnote 11; the final figure will be disclosed in the next weeks.]

4.2 Permutation Test: Pure Granularity and No Place Effects

Section 4.1 has reiterated the fact that locations differ substantially in their observed average productivity. However, this result could spuriously reflect dispersion in idiosyncratic plant productivities rather than systematically different true place effects. As a starting point, we assess whether we can rule out the extreme hypothesis that there is no variation in place effects.

Implementation We randomly relabel plant's location l, preserving the plant-count distribution across locations $l \in L$ in each 4-digit industry. Since we draw without replacement, all plants are used exactly once per randomization. We generate 1,000 randomized economies indexed by $r \in R = \{1, ...1000\}$.

Results The average variances for the random economies are 0.016 for the place effects and 0.043 for the industry-place effects, compared to 0.024 and 0.048 for the empirical ones discussed above. That is, exactly due to granularity, imposing this hypothetical data-generating process of assigning plants' productivities randomly over places already generates tremendous – purely spurious – variation in place averages. Consistent with the role of granularity captured by Equation 5, the smaller cells in the industry-place effects generate even larger variances.

For illustration, we also visualize the distribution of place effects in the randomized economy in form of the dot-dashed red line, along with the empirical distribution (solid black line), in Figure 1. [In the subsequent version, we will construct the benchmark by taking average densities over all permutations. Average densities will be disclosed, this draft follows the method in Footnote 11, by rescaling the variance of the empirical curve.] The closer the two lines, the more similar the random economies are to the empirical distribution, and the more granularity alone could account for the observed dispersion.

In Figure 2, we present the distribution of variances of the 1,000 random economies, the mean of which we reported above. This distribution is the nonparametric sampling distribution of the variance of place averages under the assumption of no place effects. The vertical dashed line denotes the level of the empirical variance. For place effects, Panel (a) clarifies that the empirical variance is within (in fact, above) the top 1 of the 1000 permutation values given the sampling distribution; the place-industry effects in Panel (b) is at the 963th observation of the 1000 permutation values, implying a one sided p-value of 0.037. That is, the empirical variance is statistically more dispersed than would be expected from a null hypothesis of no place effects, even though such economies would generate substantial purely spurious variance.

We conclude that place effects exist and must contribute to the observed variance of place averages of productivity. But to obtain an estimate of the variance of true place effects, we cannot simply, for instance, subtract the raw variance from the counterfactual mean variance of the the random economies as, in that exercise, plants' a_p would take with them their true place effects τ_l. Next, we implement our split-sample correction to quantify the variance of true place effects.

4.3 Bias Correction of Variance: Split Samples

We implement our split-sample procedure in order to remove the granularity bias, as described in Section 2.4.

Implementation While in expectation the covariance is an unbiased estimator of the variance, granularity may still imply substantial error in one given random sample split. To eliminate this concern, we implement 1,000 random sample splits, and extract the resulting distribution of covariances. We report the mean and the confidence interval of these covariances. [The current draft reports 1 draw; the future version will report on the currently undisclosed 1,000 draws after the Census RDCs reopen post-Covid.]

Results The measurement error correction dramatically reduces the variance of productivity across US regions, by two-thirds for place effects, from 2.4% to 0.8%. For industry-specific place effects, the variance is reduced from 4.8% to 1.3%. Figure 1 illustrates the effect of the bias-correction on the distribution of location and location-industry effects as a blue dashed line, drawing on a simple mean-preserving linear transformation of the raw distribution to match the measurement-error adjusted variance.[11] Hence, on the one hand, around two-thirds of the cross-regional variation in productivity simply reflects the bias arising from the idiosyncratic variation in plant-level TFP. On the other hand, the remaining variation constitutes the smaller but still considerable variance in true place effects on productivity, consistent with results from our permutation tests.

While the small industry-MSA cell size suggested considerable room for granularity bias, perhaps more surprising is the similar bias in the variance of MSA effects. Appealing to Equation (5), this result likely reflects large plant-level, idiosyncratic within-location variance.

[While the numbers cited above are reflect the final specification, the visualization in Panel (b) is currently based on one ignoring year-industry demeaning; the final figure providing averages of the distributions across the 1,000 sample splits will be disclosed once Census RDCs reopen post-Covid.]

Visualization: Binned Scatter Plot We visualize the measurement error correction in Figure 3; Panel (a) reports on the place effects, Panel (b) on the industry-place effects. The panels are binned scatter plots, juxtaposing, for each MSA, its pair of place effects computed separately on the basis of the samples split in half within each MSA.[12] The x-axis describes the means of each bin from split sample A. On the x-axis we therefore trace out the dispersion in place effect in the raw data (which as argued above is more dispersed for a half sample). The y-axis plots the corresponding place effect for the MSA's place effect implied by its split sample B excluded from the x-axis. Intuitively, if there were no relationship whatsoever between the two samples A and B, the graph would trace out a line with slope zero through the origin. In the absence of idiosyncratic effects, the place effects across the subsamples within the location should be identical, so that the scatter plots would trace out the 45 degree line. We include both these lines as benchmarks.

We also include the linear regression line implied by the underlying MSA-level (rather than binned) data. We estimate a slope of 0.216 for place effects, and a slope of 0.126 for industry-specific place effects.[13] The linear regression coefficient is of interest as it corresponds to $\gamma = \frac{\mathsf{Cov}(\widehat{\xi}_l^{S^{s'}}, \widehat{\xi}_l^{S^s})}{\mathsf{Var}(\widehat{\xi}_l^{S^s})}$. Since $\mathsf{Cov}(\widehat{\xi}_l^{S^A}, \widehat{\xi}_l^{S^B})$ is our measurement-error corrected

[11] That is, we construct $x' = a + bx$ and $f_{x'} = f((x-a)/b)/b$. We set $b = \frac{\mathsf{Var}(x')}{\mathsf{Var}(x)}$ to match the desired variance of the transformed distribution, and $a = (1-b)\mathrm{E}[x]$ to preserve the mean.

[12] The current binned scatter plot uses a split sample of only establishments older than 5 years. Pending reopening of Census RDCs, we will release the same figure for the full sample.

[13] The standard errors of those slope estimates are 0.039 and 0.014, respectively; we do not interpret this standard error and ignore that the fixed effects are generated regressors.

estimate of the variance $\mathsf{Var}(\widehat{\xi}_l^S)$, the regression coefficient represents the share of the raw variance (of the half sample depicted on the x-axis) that survives bias correction. These positive slopes therefore confirm the small, yet attenuated, presence of true place effects. Read literally, they imply 87.4% and 78.4% of the raw variance in place and place-industry effects are due to granularity bias. This fraction is smaller that comparing the split-sample covariance with the population variance, which has exactly twice as many firms per cell, so splitting the sample in half in each MSA (ignoring odd numbers) naturally doubles the granularity bias term in Equation (5) for any original sample S and split samples $s \in \{A, B\}$, i.e. $\frac{1}{L}\sum_{l \in L} \frac{N_{S_l}}{N_{S_l^s}} \frac{\sigma_l(u)^2}{N^{S_l}} = 2 \times \frac{1}{L}\sum_{l \in L} \frac{\sigma_l(u)^2}{N^{S_l}}$ since $\frac{N_{S_l}}{N_{S_l^s}} = 2 \forall l \in L$.[14]

5 Application: The Productivity of New Plants

We now separately measure location effects for the subset of new plants. *A priori*, there is considerable room for the place effects for new plants to differ from the pooled averages constructed in Section 4. The pooled effects are dominated by older and larger incumbent plants. Empirically, in 2012, plants older than 5 years made up 78% of manufacturing plants, and 91% of total employment in the manufacturing sector (US Census Business Dynamics Statistics, 2014 Release). On the theoretical side, in models of embodied technological change (as in, e.g., Sakellaris and Wilson, 2004) new projects reflect the frontier technology while incumbent, old projects reflect legacy technologies. Some models (Duranton and Puga, 2001, e.g.,) specifically predict that some cities are better environments for entrepreneurship than others.

Strategy We partition the plants into new plants (superscript Y for "young", not N, which we use for counts) aged five years and younger, and old plants (O) aged six years and older. The five-year age cutoff is chosen as the Census occurs every five years. Because TFP may be more noisily measured for new plants, Appendix Table A.1 presents the results for log value added per worker.

Dispersion Figure 4 plots, in thick black lines, the distribution of raw location effects for new plants, $\widehat{\xi}_l^Y$, in Panel (a) and for location-industry effects, $\widehat{\tau}_{l,i}^Y$ in Panel (b). For both

[14] Consistent with our main finding that the granularity bias term dominates the place effects, we indeed find that the split sample raw variance is 150-200% as large as the population variance for both the place effects ($\mathsf{Var}(\widehat{\xi}^{S_l^2}) = 0.036$ vs. $\mathsf{Var}(\widehat{\xi}^{S_l}) = 0.024$) and the industry-place effects ($\mathsf{Var}(\widehat{\tau}_{l,i}^{S_l^s}) = 0.100$ vs. $\mathsf{Var}(\widehat{\tau}_{l,i}^{S}) = 0.048$). With these benchmark of a larger raw variance of the split sample, the bias corrections on the basis of the covariances of the split samples, $\mathsf{Cov}(\widehat{\xi}_l^{S^A}, \widehat{\xi}_l^{S^B})$ and $\mathsf{Cov}(\widehat{\tau}_{l,i}^{S^A}, \widehat{\tau}_{l,i}^{S^B})$, would only leave around around 1/6 or 1/8 rather than 1/3 of the initial raw variance intact.

definitions, place effects of new plants are around twice as dispersed at $\mathsf{Var}(\widehat{\xi}_l^Y) = 0.060$ and $\mathsf{Var}(\widehat{\tau}_{l,i}^Y) = 0.130$. From the density plots, the top 5% of place effects appear 100% higher than the bottom 5-percentile; that p95/p5 premium is nearly 200% for industry-place effects. Appendix Table A.1 provides the additional statistics of this analysis.

The figure also plots, in thin black lines, the distributions of old plants' place effects, which mirror that of the pooled averages, as expected given their quantitative dominance. The raw variances are $\mathsf{Var}(\widehat{\xi}_l^O) = 0.031$ and $\mathsf{Var}(\widehat{\tau}_{l,i}^O) = 0.058$ (the slightly higher variances than the pooled effects from Section 1 are likely due to smaller cell sizes).

Some of the higher variance of the new plants' place effects may simply reflect heightened granularity bias, due to smaller populations and potentially higher idiosyncratic dispersion relative to the pooled specifications. Indeed, the bias-corrected variances of the new plants drop dramatically to $\mathsf{Cov}(\widehat{\xi}_l^{Y,A}, \widehat{\xi}_l^{Y,B}) = 0.008$ for location effects and $\mathsf{Cov}(\widehat{\tau}_{l,i}^{Y,B}, \widehat{\tau}_{l,i}^{Y,B}) = 0.014$ for location-industry effects. These corrections entail much larger reductions from the raw variances than for the pooled place effects in Section 4.[15]

We also bias-correct the raw variances with the split-sample strategy. We find $\mathsf{Cov}(\widehat{\xi}_l^{O,A}, \widehat{\xi}_l^{O,B}) = 0.010$ and for location effects and $\mathsf{Cov}(\widehat{\tau}_{l,i}^{O,B}, \widehat{\tau}_{l,i}^{O,B}) = 0.013$ for location-industry effects, again close to the pooled effects from Section 4 of 0.008 and 0.013. We conclude that the higher dispersion of new plants' place effects is due to heightened granularity bias.

Juxtaposing Place Effects of New Plants and Old Plants With the bias-corrected dispersion among place effects for young plants strikingly similar to that of incumbents, a natural question is whether they are correlated across these age groups. For example, a nursery cities view Duranton and Puga (2001) would permit some cities to be particularly suitable for entrepreneurship, in ways that need not carry over to incumbent, large and old production units. Alternatively, place-based productivity differences could be entirely cohort-specific.

To investigate this question, we juxtapose the new-plants place effect of a given MSA with the place effect of old plants only, a leave-out approach that avoids mechanical correlation between the axes. As one benchmark, consider a slope of one: place effects would show up for both young and old projects identically. As another other benchmark, a slope of zero would indicate no relationship between new and old plants' respective place effects.

Figure 4 Panels (b) and (c) present these scatter plots for the location and location-

[15] We also implement the permutation tests for the young (and old). Location effects for the young remain significantly dispersed over a random location benchmark ($p = 0.006$). Interestingly, for MSA-industry effects we find $p = 0.375$, perhaps indicating very small cell sizes. A future version will draw out the CIs of covariances from 1,000 sample splits, rather than the current one split.

industry effects respectively. The red line represents the linear slope γ^{RF}, which is the reduced-form effect in the IV interpretation of the split-sample method we describe below. The slopes reveal a seemingly small elasticity of 0.08 and 0.10 for location and location-industry effects of new-plants to old-plants place effects.

However, in light of the noisily estimated place effects, the unity benchmark is an inappropriate target: just as the raw variance greatly overstates the true place effect dispersion among the old, we expect attenuation bias in this elasticity estimate.

We construct a bias-corrected adjustment of the benchmark by estimating a first stage of the split samples in an IV interpretation, from a regression of the place effects a random half sample of old plants $S^{O,B}$ (y-axis) on those on the other sample of old plants $S^{O,A}$ (x-axis). The blue line plots the resulting regression slope γ^{FS} of the first stage. This strategy lowers benchmark from a coefficient of one to this first stage coefficient pair of 0.216 for place effects and 0.126 for industry-specific place effects.[16]

This first stage is analogous to the visualization of the overall bias correction in the full sample, which we had depicted in the form of a scatter plot in Figure 3, except that we now conduct it among the old plants. [Recall that the currently plotted Figure 3 is based on old plants; the version pooling all plants will be disclosed once the Census RDCs reopen post-Covid.]

With this bias-corrected benchmark, the new plants appear to share a larger share of the place effects of the old, but the relationship remains far from identical. We also report the IV effect $\gamma^{\text{IV}} = \gamma^{\text{RF}}/\gamma^{\text{FS}}$ drawing on first stage effect γ^{FS}, which is 0.264 for place effects and 0.600 for industry-place effects. Intuitively, the IV effect measures the distance to that corrected benchmark of full sharing.

We conclude that while spatial dispersion of place effects are similar for old and young plants, the two sets of place effects are indeed distinct to a substantial degree, between one fifth for place effects and three fifths for industry-specific place effects.

6 Application: The Countries of Europe

We close by applying our method to NUTS-2 regions of European countries, drawing on Bureau van Dijk data.

Figure 5 presents the results for the regional dispersion in TFP. The figure recapitulates the US Census based findings as the very top entry. For each country, we report three statistics of dispersion: (i) the raw variance (solid red circle), (ii) the mean variance implied by 1,000 random allocations of plants across places (hollow red diamond)

[16] F-statistics for the first stage regressions in Panels (a) and (b) are 45 and 15, respectively.

along with the 2.5% and 97.5% confidence intervals (dashed red line) taking from the sampling distributions given by the 1,000 randomization), and (iii) the mean covariance – the measurement-error-adjusted estimate of the variance – of the 1,000 randomly split samples (blue triangle) along with 2.5% and 97.5% confidence intervals (solid blue line).

Consistent with the US findings, the European case studies exhibit large variation of measured productivity place effects, although raw estimates differ by country. Yet, across most countries, random allocations would have yielded similar dispersion, such that the empirical raw variance is well within the distribution of the permutations for 8 countries, and the null can only be rejected in 7. Similar to our findings in the US data, the granularity bias adjustment is usually statistically different from the variance but varies considerably by country – from less than 15% for Portugal, to 100% for the United Kingdom and Austria. Unlike the US, for no European country is the variance of industry-specific location effects statistically and economically different from the permuted variance, and all corrected location-industry variances are close to zero. These differences from the US data may be due to the data limitations discussed in Section 3.

7 Conclusion

We measure dispersion in productivity across locations. While we confirm the canonical high spatial variation in the raw data, we find that the majority, around two thirds, of this raw variation is an artifact of cells of finite plants with high idiosyncratic variation. This pattern holds in the US as well as in the 15 European countries we additionally study. Furthermore, we uncover substantially more variation in location effects measured from new plants. All of this excess variance is due to heightened granularity bias.

We close by reiterating that our measurement-error corrected numbers are model-agnostic and reduced-form. They are consistent with a wide array of theories of place. They may reflect a combination of causal effects of place on productivity (such as agglomeration forces), as well as sorting. Moreover, our method does not correct for even more fundamental sources of measurement error, for example arising from location-specific production function misspecification or mismeasurement of inputs or prices. Hence, the remaining true dispersion in place effects may be even lower than our bias-corrected estimates. Overall, our results suggest a substantially reduced upper bound for the potential productivity impact of the actual underlying spatial forces, to the degree that they show up through location-specific expected values of plant productivity.

References

Bils, Mark, Peter Klenow, and Cian Ruane. 2020. "Misallocation or Mismeasurement?" *NBER Working Paper* .

Boeri, Tito, Andrea Ichino, Enrico Moretti, and Johanna Posch. 2019. "Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces." *CEPR Discussion Paper* .

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte. 2018. "The Impact of Regional and Sectoral Productivity Changes on the US Economy." *The Review of Economic Studies* 85 (4):2042–2096.

Davis, Donald and Jonathan Dingel. 2020. "The Comparative Advantage of Cities." *Journal of International Economics* 123:103291.

Dingel, Jonathan and Felix Tintelnot. 2020. "Spatial Economics for Granular Settings." *NBER Working Paper* .

Drenik, Andres, Simon Jäger, Pascuel Plotkin, and Benjamin Schoefer. 2020. "Paying Outsourced Labor: Direct Evidence from Linked Temp Agency-Worker-Client Data." *NBER Working Paper* .

Duranton, Gilles and Diego Puga. 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products." *American Economic Review* 91 (5):1454–1477.

Ellison, Glenn and Edward Glaeser. 1997. "Geographic Concentration in US Manufacturing Industries: a Dartboard Approach." *Journal of Political Economy* 105 (5):889–927.

Fajgelbaum, Pablo and Cecile Gaubert. 2018. "Optimal Spatial Policies, Geography and Sorting." *NBER Working Paper* .

Fajgelbaum, Pablo, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar. 2019. "State Taxes and Spatial Misallocation." *The Review of Economic Studies* 86 (1):333–376.

Fort, Teresa and Shawn Klimek. 2018. "The Effects of Industry Classification Changes on US Employment Composition." *Working Paper* .

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98 (1):394–425.

Gaubert, Cecile. 2018. "Firm Sorting and Agglomeration." *American Economic Review* 108 (11):3117–53.

Gaubert, Cecile and Oleg Itskhoki. forthcoming. "Granular Comparative Advantage." *Journal of Political Economy* .

Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card. 2018. "Assortative Matching or Exclusionary Hiring? The Impact of Firm Policies on Racial Wage Differences in Brazil." *NBER Working Paper* .

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. "Capital Allocation and Productivity in South Europe." *The Quarterly Journal of Economics* 132 (4):1915–1967.

Hornbeck, Richard and Enrico Moretti. 2018. "Who Benefits from Productivity Growth? Direct and Indirect Effects of Local TFP Growth on Wages, Rents, and Inequality." *NBER Working Paper* .

Hsieh, Chang-Tai and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11 (2):1–39.

Jäger, Simon, Benjamin Schoefer, and Jörg Heining. 2019. "Labor in the Boardroom." *NBER Working Paper w26519* .

Moretti, Enrico. 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt.

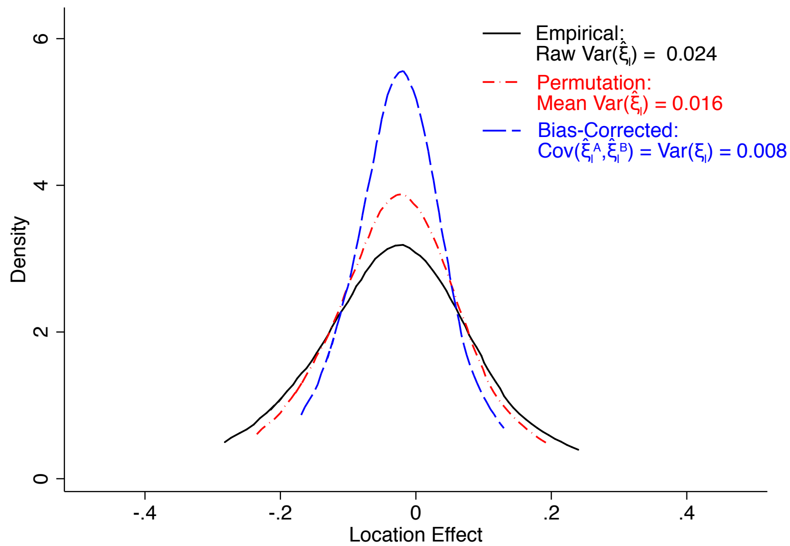
Sakellaris, Plutarchos and Daniel J Wilson. 2004. "Quantifying Embodied Technological Change." *Review of Economic Dynamics* 7 (1):1–26.

Silver, David. 2020. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *Working Paper* .

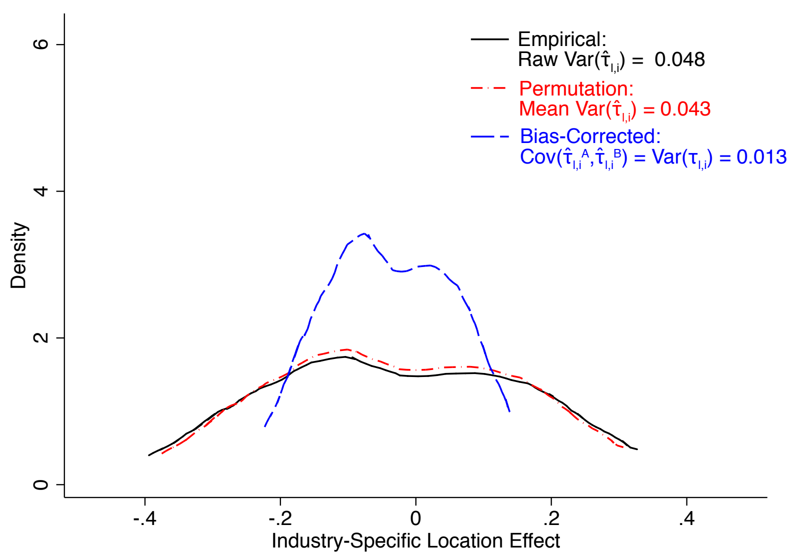
Figures

Figure 1: Place Effects on TFP Across US MSAs: Empirical, 1,000 Random Permutations of Plants Across Places, and Corrected for Measurement Error (Granularity Bias)

(a) Location Effects



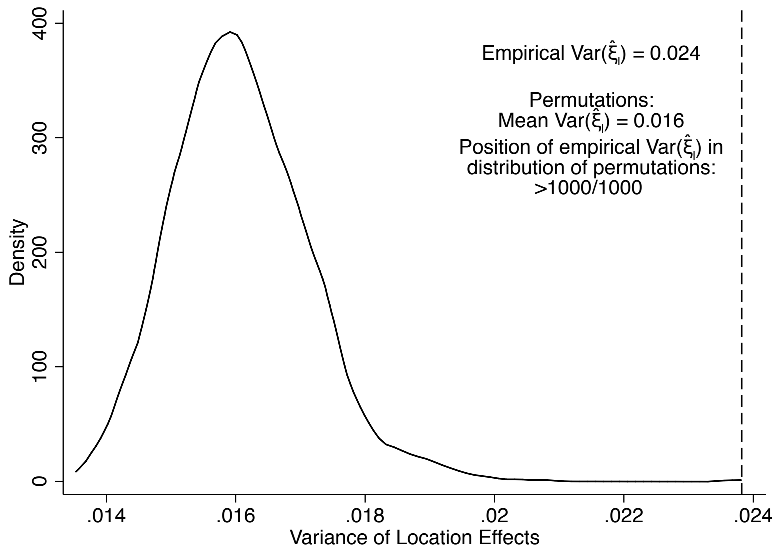
(b) Location-Industry Effects



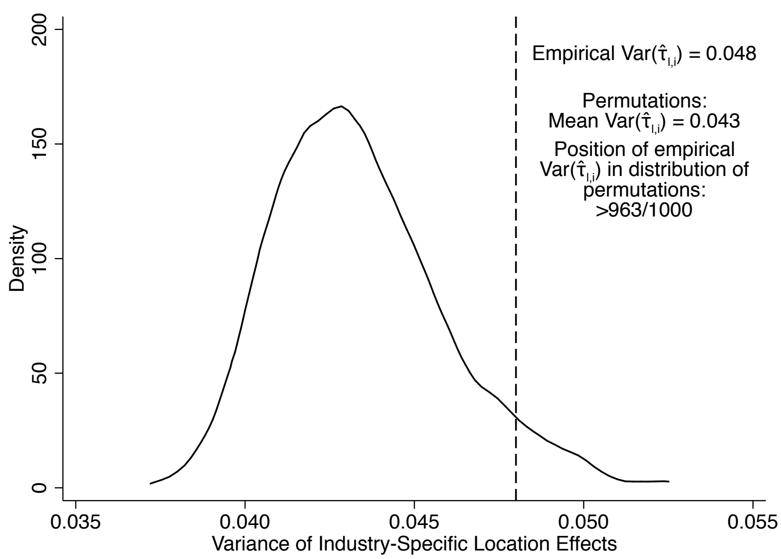
Note: The panels are kernel density plots representing the distribution of estimated place effects (Panel (a)) and place-industry effects (Panel (b)). The solid black line plots the distribution of the empirical effects. The dash-dotted red line shows the distribution of place effects from a permutation tests (plotting the average densities). The blue dashed line illustrates the biased-corrected distribution. It does so by applying a mean-preserving variance-adjustment using a linear transformation of the original distribution as described in Footnote 11. The locations are US MSAs; the industry definition is 4-digit NAICS. [Currently, while the permutation line is printed to exhibit the average variance over all permutations, the permutations-based density is constructed using the rescaling procedure mirroring that of the bias-corrected line. When the Census RDC reopens, the permutations-based line will be replaced by one capturing the actual average density.] The data set is the US Census of Manufacturers, and includes industry-location cells with at least 2 young and 3 old plants. Appendix Table A.1 reports additional statistics and robustness.

Figure 2: Permutation Test: Sampling Distribution of Variances of Place Effects From 1,000 Random Assignments of Plants Over Places

(a) Location Effects



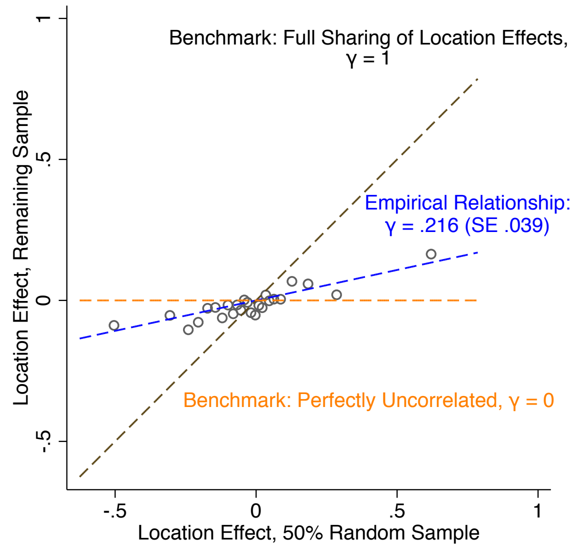
(b) Location-Industry Effects



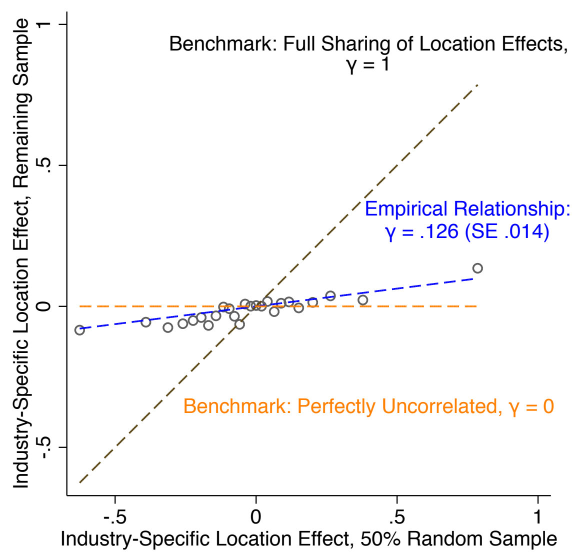
Note: The panels report kernel density plots corresponding to the sampling distribution of 1,000 economies' variances of estimated place effects (Panel (a)) and place-industry effects (Panel (b)). The randomization procedure reassigns MSA IDs over the empirical plant observations within an industry, preserving the industry-MSA plant count distribution. The vertical dashed line denotes the empirical variance. Companion Appendix Table A.1 reports additional statistics and robustness. [This figure will be updated once the Census RDCs reopen.]

Figure 3: Illustration of Method: Split-Sample Correction of Raw Variance Removing Granularity Bias

(a) Location Effects



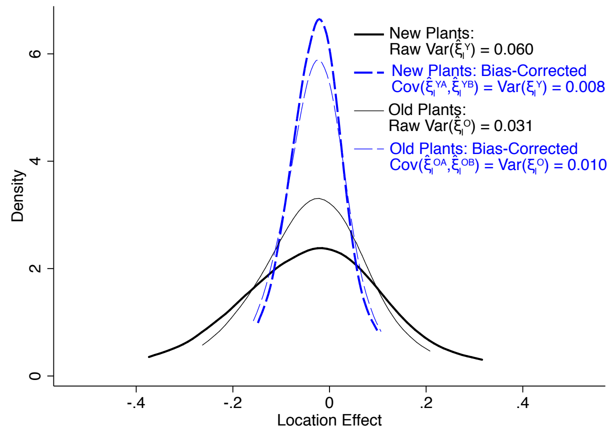
(b) Location-Industry Effects



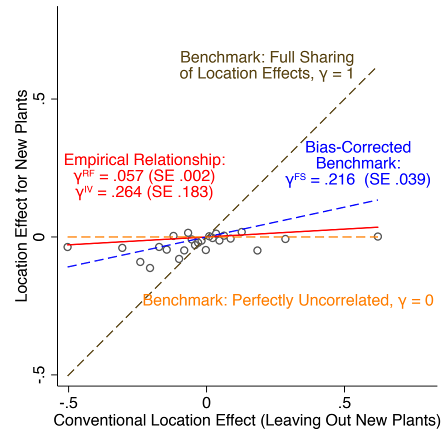
Note: The panels are scatter plots juxtaposing, along the common location or location-industry ID, the estimated place effects from one split sample on place effects of the other split sample. We also plot two benchmarks. $\gamma = 0$ represents the scenario of no place effects whatsoever, i.e. no relationship between place effects of the split samples; $\gamma = 1$ represents the scenario in which place effects feature no attenuation bias from measurement error such as granularity. The blue line traces out the linear slope from the regression of the y-axis effects (one split sample) against those of their x-axis neighbor (other split sample). Since the underlying univariate regression coefficient represents the covariance of the variables on the two axes divided by the variance of the x-axis variable, and since the split-sample covariance is the bias-corrected estimator of the variance, this coefficient also represents the share of the variance surviving the bias correction. Note that the denominator, the raw variance of the split-sample, is larger exactly due to heightened granularity bias going along with the halving of the sample size, so this ratio is around only half the size of the corresponding ratio taking with the raw variance of the full sample, which forms our preferred and more conservative statistic. [Currently, the plots are based on a sample of plants older than 5 years (which however make up more than 90% of the sample). Pending Census RDC reopening, the age restriction will be removed.]

Figure 4: Place Effects of New vs. Old Plants

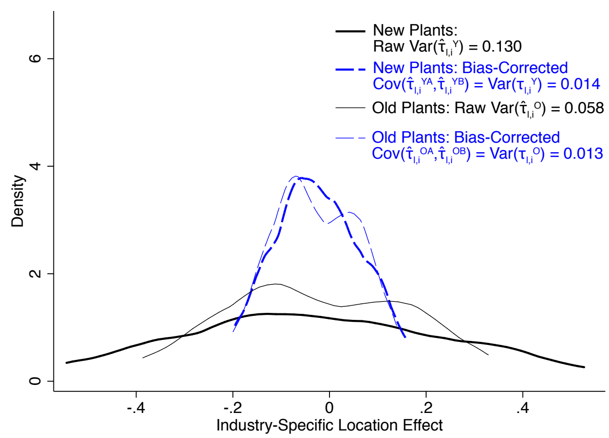
(a) Location Effects: Distributions



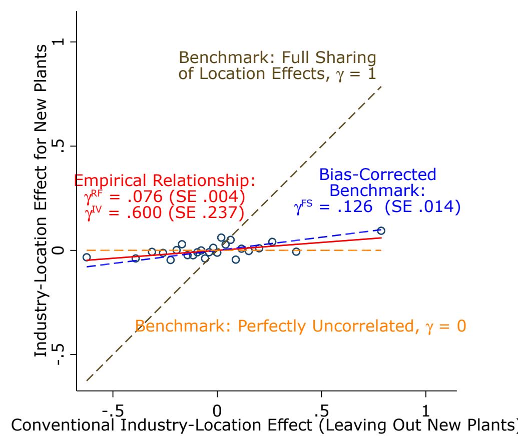
(c) Location Effects: Scatter Plots



(b) Location-Industry Effects: Distributions



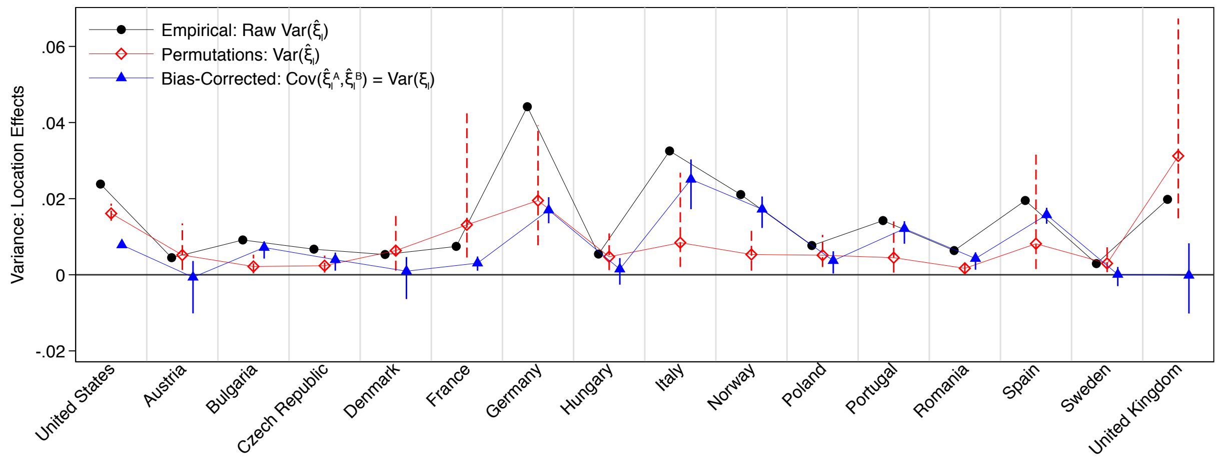
(d) Location-Industry Effects: Scatter Plots



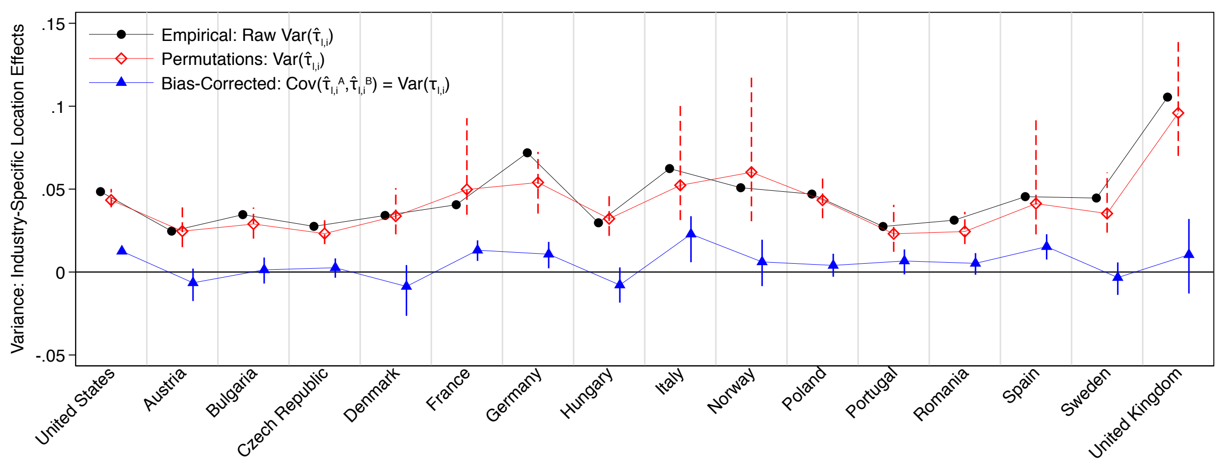
Note: The figure presents place effects for new and old plants. The left panels are kernel density plots representing the distribution of estimated place effects (Panel (a)) and place-industry effects (Panel (b)). The solid black line plots the empirical distribution of the raw effects. The blue dashed line illustrates the biased-corrected distribution. It does so by applying a mean-preserving variance-adjustment using a linear transformation of the original distribution as described in Footnote 11. The thick lines refer to place effects of new plants (5 and younger); the thin lines refer to the effects of remaining plants, i.e. those older than 5 years. Appendix Table A.1 reports the underlying numbers and replicates the results for log value added per worker rather than TFP depicted here. The right panels show binned scatter plots of estimated place effects (Panel (c)) and place-industry effects (Panel (d)) among young firms (aged 5 and younger) on the y-axis, plotted against their older neighbors' place effects in the same place on the x-axis. The red line represents the linear slope γ^{RF}, which is the reduced-form effect in the IV interpretation of the split-sample method. We also report the IV effect $\gamma^{IV} = \gamma^{RF}/\gamma^{FS}$ drawing on first stage effect γ^{FS}, which we describe next. The blue line plots the regression slope γ^{FS} of the first stage, which is obtained by regressing the half sample of old plants' place effects on the y-axis with the place effects estimated in the complementary set of old plants on the x-axis. F-statistics for the first stage regressions in Panels (a) and (b) are 45 and 15, respectively. We also plot two benchmarks. $\gamma = 0$ represents no relationship between place effects of old and new plants; $\gamma = 1$ represents the naive effect assuming no attenuation bias from measurement error such as granularity. The appropriate comparison for full sharing of place effects is the first-stage slope. Intuitively, the IV effect measures the distance to that corrected benchmark of full sharing.

Figure 5: Place Effects by Country: United States and European Countries

(a) Location-Specific Effects



(b) Location-Industry-Specific Effects



Note: The panels report, for each country, raw variance (solid black dot), mean and 95% CI for variance of permutations (hollow red diamond and dotted line, respectively) , and mean and 95% CI of bias-corrected variance (solid blue triangle and line, respectively). Panel (a) reports the location (NUTS2) effects and Panel (b) reports the location (NUTS2) x industry (NACE 2-digit) effects for European countries; the top reiterates the US numbers using the MSA and 4-digit NAICS cell definitions. The place effects for the additional, European countries are measured using Bureau van Dijk (BvD) firm-level data, in which we construct TFP using fixed tangible assets as the proxy for the firm's capital stock, and compute industry-specific capital shares assuming a standard Cobb-Douglas production function. The firms are from the most recent cross-section for each country; to emulate a census and to account for incomplete coverage, we supplement this wave with additional firms observed in previous BvD waves, so that for each firm, we use the most recent and exactly one observation. Appendix Table A.2 details the statistics printed here. Appendix Figure A.2 replicates this figure but estimates place effects on the basis of firms' log value added per worker rather than TFP.

Appendix

A Additional Tables

Table A.1: Robustness and Additional Statistics: Productivity Dispersion Across Locations and Location-Industries

	All Plants					New Plants		Old Plants	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	TFPr	$\ln \frac{\text{Value Added}}{\text{Emp}}$	Unweighted	6-digit	≥ 10 plants	TFPr	$\ln \frac{\text{Value Added}}{\text{Emp}}$	TFPr	$\ln \frac{\text{Value Added}}{\text{Emp}}$
Panel A: Variance of Place Effects									
Raw variance	0.024	0.063	0.012	0.019	0.017	0.060	0.139	0.031	0.081
Mean of perm's	0.016	0.042	(pending RDC reopening)			0.062	0.162	0.020	0.051
SD of perm's	0.001	0.002	(pending RDC reopening)			0.005	0.009	0.001	0.003
p-value	0.000	0.000	(pending RDC reopening)			0.006	0.088	0.000	0.000
Split-sample cov.	0.008	0.025	0.005	0.003	0.006	0.008	0.015	0.010	0.028
Panel B: Variance of Place-Industry Effects									
Raw variance	0.048	0.115	0.023	0.034	0.029	0.130	0.284	0.058	0.139
Mean of perm's	0.043	0.104	(pending RDC reopening)			0.128	0.303	0.052	0.121
SD of perm's	0.003	0.005	(pending RDC reopening)			0.008	0.013	0.003	0.006
p-value	0.037	0.034	(pending RDC reopening)			0.375	0.940	0.046	0.009
Split-sample cov.	0.013	0.031	0.007	0.003	0.008	0.014	0.030	0.013	0.034
N: MSAs	(pending RDC reopening)								
N: industry-MSAs	(pending RDC reopening)								
N: plants	(pending RDC reopening)								

Note: The table reports variances of location and location-industry productivity effects in US manufacturing using the Census of Manufactures. Specification (1) is our preferred sample and specification. All other columns report specifications that each change one aspect of specification (1), as follows: Specification (2) uses plant-level log value added per worker rather than plant-level TFP. Specification (3) is unweighted rather than using plant-level employment weights. Specification (4) uses 6-digit rather than 4-digit NAICS industry levels. Specification (5) applies a minimum of 10 plants per cells rather than 5 (specifically 6 (rather than 3) and 4 (rather than 2) plants ages over 5 years and 5 years or fewer, respectively). Specifications (6)-(9) separately replicate specifications (1) and (2) for new plants (aged 5 and younger) and old plants (older than 5 years), for TFP (columns (6) and (8)) and log value added per worker (columns (7) and (9)).

Table A.2: Europe: The Dispersion of Productivity (TFP) Across Locations, by Country

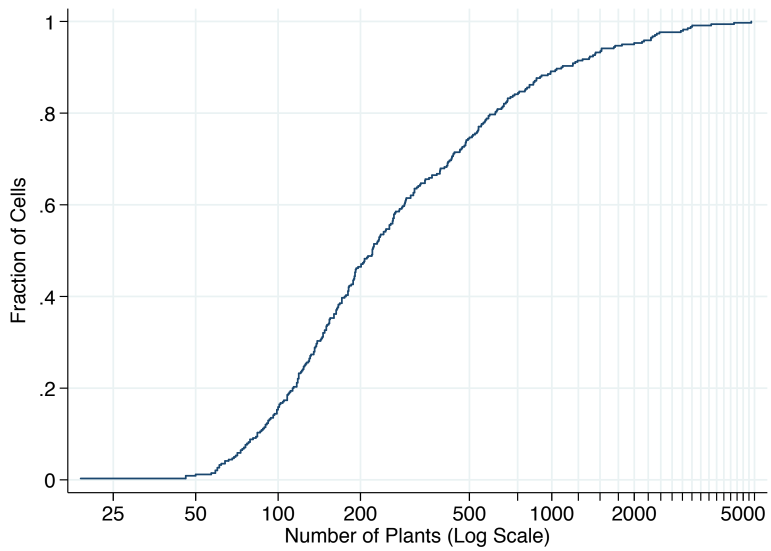
Country	Count		Raw	Permutations			Bias-Corr.
	Cells	Firms	Var	Mean Var	95% CI	p-Value	Var
Panel A: Location Effects							
Austria	9	1,540	0.004	0.005	(0.001, 0.013)	0.506	-0.001
Bulgaria	6	14,601	0.009	0.002	(0.000, 0.005)	0.001	0.007
Czech R.	8	13,067	0.007	0.002	(0.001, 0.005)	0.006	0.004
Denmark	5	5,805	0.005	0.013	(0.005, 0.043)	0.538	0.001
France	26	31,895	0.007	0.031	(0.015, 0.067)	0.773	0.003
Germany	38	12,790	0.044	0.006	(0.001, 0.016)	0.011	0.017
Hungary	8	2,672	0.005	0.008	(0.002, 0.027)	0.317	0.001
Italy	21	116,845	0.033	0.005	(0.001, 0.012)	0.017	0.025
Norway	7	5,158	0.021	0.005	(0.002, 0.010)	0.000	0.017
Poland	17	6,430	0.008	0.004	(0.001, 0.014)	0.122	0.004
Portugal	7	25,969	0.014	0.002	(0.000, 0.004)	0.025	0.012
Romania	8	21,254	0.006	0.008	(0.002, 0.032)	0.000	0.004
Spain	17	60,312	0.020	0.020	(0.008, 0.039)	0.120	0.016
Sweden	8	13,863	0.003	0.003	(0.001, 0.008)	0.415	0.000
UK	42	9,468	0.020	0.005	(0.001, 0.011)	0.855	0.000
Panel B: Location-Industry Effects							
Austria	100	1,540	0.025	0.025	(0.015, 0.039)	0.506	-0.006
Bulgaria	124	14,601	0.035	0.029	(0.020, 0.039)	0.001	0.001
Czech R.	162	13,067	0.028	0.023	(0.017, 0.031)	0.006	0.003
Denmark	97	5,805	0.034	0.050	(0.035, 0.093)	0.538	-0.009
France	440	31,895	0.041	0.096	(0.070, 0.139)	0.773	0.013
Germany	564	12,790	0.072	0.034	(0.023, 0.051)	0.011	0.011
Hungary	129	2,672	0.03	0.052	(0.031, 0.102)	0.317	-0.008
Italy	428	116,845	0.062	0.060	(0.031, 0.117)	0.017	0.023
Norway	125	5,158	0.051	0.043	(0.032, 0.058)	0.000	0.006
Poland	252	6,430	0.047	0.023	(0.012, 0.041)	0.122	0.004
Portugal	109	25,969	0.027	0.024	(0.017, 0.036)	0.025	0.007
Romania	167	21,254	0.031	0.041	(0.023, 0.092)	0.000	0.005
Spain	341	60,312	0.045	0.054	(0.035, 0.072)	0.120	0.015
Sweden	156	13,863	0.045	0.035	(0.024, 0.060)	0.415	-0.003
UK	508	9,468	0.105	0.032	(0.022, 0.047)	0.855	0.010

Note: Cell definitions are NUTS-2 regions in Panel A and NUTS-2 region x NACE 2 industry in Panel B.

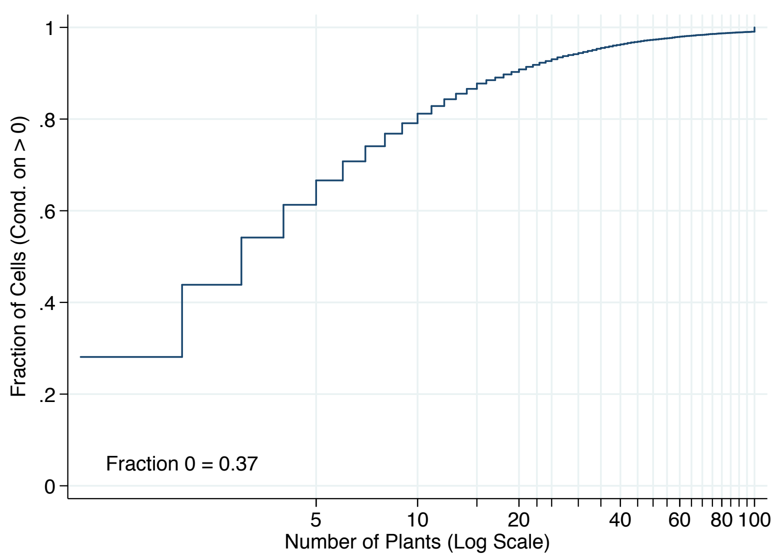
B Additional Figures

Figure A.1: Plant Counts in US Manufacturing by Cell Definition

(a) MSA



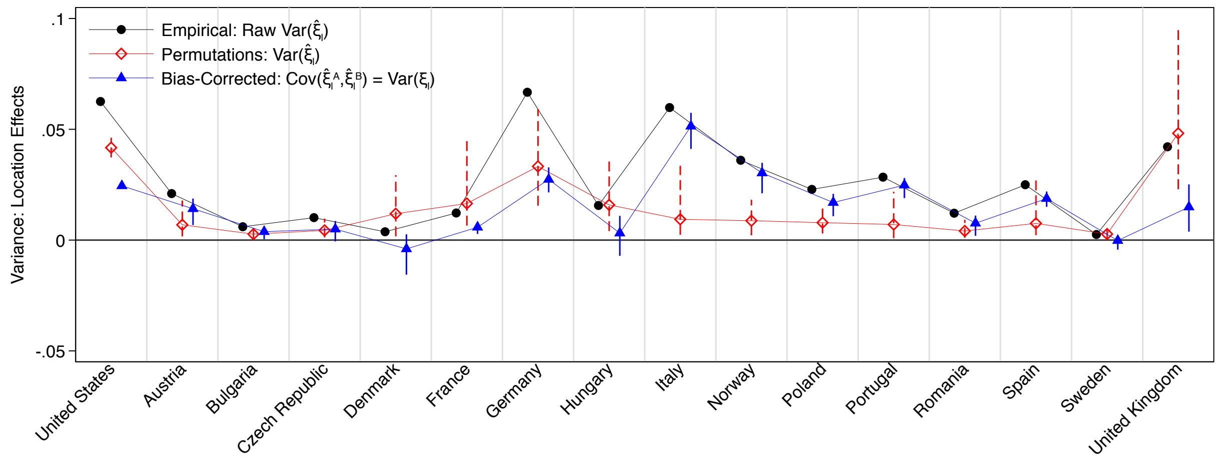
(b) Industry-MSA



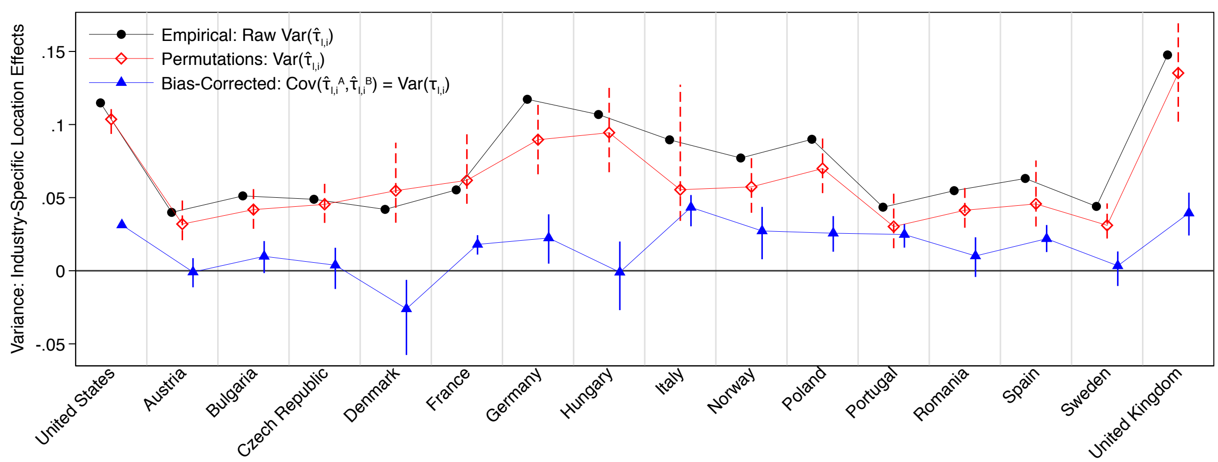
Note: The figure presents plant counts in US MSAs (Panel (a)) and in MSA x 4-digit NAICS cells (Panel (b)), in 2016 in the manufacturing sector. Panel (b) is truncated at 100 plants. 37% of location-industry cells have zero plants. The source is the 2016 County Business Patterns.

Figure A.2: Place Effects by Country: United States and European Countries, Value Added

(a) Location-Specific Effects



(b) Place-Industry-Specific Effects



Note: The figure replicates Figure 5 but estimates place effects on the basis of firms' log value added per worker rather than TFP.