Productivity, Place, and Plants

CEPR/University of Milan Annual Symposium In Labour Economics (ESSLE)
June 2021

Benjamin Schoefer
UC Berkeley

Oren Ziv
Michigan State University

Connect two phenomena:

Urban economics: places di er in productivity (various reasons/mechanisms)

Micro productivity literature: plants exhibit tremendous idiosyncratic heterogeneity

Q: How much do plants differ in productivity across places for reasons that are systematically related to their current location? Which role for "luck of the draw" of individual plants?

Goal: Measure dispersion in (manufacturing) productivity (TFPr, labor prod.) across US cities (MSAs) and isolate cross-regional variance in true place effects

- Strip out bias from idiosyncratic plant-level heterogeneity (\granularity bias") from raw cross-MSA variance in prod.

Key ndings:

Large raw cross-MSA variance: avg prod in 90th pctile MSA is 60-140% higher than in 10th pctile

Large granularity bias: 2/3 to 3/4 of cross-MSA raw variance is unrelated to place

- Much smaller true place e ects: at most 1/4 - 1/3 of raw cross-MSA variance re ects true place e ects

Applications/extensions:

Robustness...

New plants - 60% pass-through of true place e ects.

Extend to w/in regional dispersion in 15 European countries

Outline

De nitions & Basic Statistics

Plant p in location $l \in L$ has productivity (log TFP) a_{pl} .

a $F_l^a(a)$ – DGP is l-specific.

Statistical def, agnostic to sources of l-dependence: sorting, agglomeration e ects, mismeasurement,...

True place effect:

$$\Psi_l = \frac{\mathbb{E}}{Z}[a_{pl} | l] = \int a dF_l^S(a)$$

$$\Rightarrow a_{pl} = \Psi_l + u_{pl}$$

Measured average productivity of count N^{S_l} plants named $p \in S_l$:

$$\hat{\Psi}_l^{S_l} = \frac{1}{N^{S_l}} \sum_{p \in S_l} a_{pl}:$$

Of course, average $\hat{\Psi}_l^{S_l}$ is an unbiased and consistent estimator of Ψ_l...

Raw Variance of Place Averages

Var of Est. Place Effects (Location Averages)

$$\underbrace{\mathrm{Var}(b_l^{S_l})}_{\text{Var of Est. Place Effects (Location Averages)}} = \mathrm{Var}\left(\frac{1}{N^{S_l}} \sum_{p \in S_l} a_{pl}\right)$$

$$= \mathrm{Var}\left(\frac{1}{N^{S_l}} \sum_{p \in S_l} [\ _l + u_{pl}]\right)$$

$$= \mathrm{Var}\left(\ _l + \frac{1}{N^{S_l}} \sum_{p \in S_l} u_{pl}\right)$$

Pitfalls: Granularity Bias

Var of Est. Place Effects (Location Averages)

$$\mathrm{Var}(b_l^{S_l}) = \mathrm{Var}\left(\frac{1}{N^{S_l}} \sum_{p \in S_l} a_{pl}\right)$$

$$= \mathrm{Var}\left(\frac{1}{N^{S_l}} \sum_{p \in S_l} [\ _l + u_{pl}]\right)$$

$$= \mathrm{Var}\left(\ _l + \frac{1}{N^{S_l}} \sum_{p \in S_l} u_{pl}\right)$$

$$= \underbrace{\mathrm{Var}(\ _l)}_{\text{Var of True Place Effects}} + \underbrace{\frac{1}{L} \sum_{l \in L} \frac{\ _l(u)^2}{N^{S_l}}}_{>0 \text{ if } N<1 \ \wedge \ (u)>0} + \underbrace{2\,\mathrm{Cov}\left(\ _l, \frac{1}{N^{S_l}} \sum_{p \in S_l} u_{pl}\right)}_{=0}$$

Bias from Granularity: Var of Sample Means

Orthogonal by Construction

Outline

Definitions & statistical basics

Raw variance

Permutation test: granularity-bias-only benchmark

Bias correction: split-sample method

Tracing the sources of granularity bias

Extension I: new plants' place effects

Extension II: within-country dispersion in 15 European countries

Construction of Place Averages: US Census of Manufactures

Industry-specific location effect:

$$b_{l(p);i(p)} = \text{Avg}[a_{plj}ji; l] \quad a_i$$

Demeaned) "Average percent premium in TFP compared to national industry average"

4-digit NAICS x MSA [Robustness: 6-digit]

2 plants per cell [Robustness: at least 10]

Overall location effect:

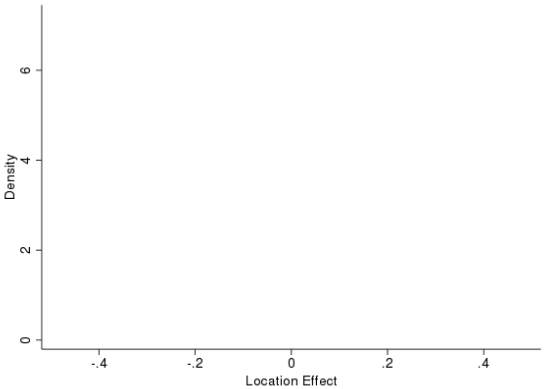
$$b_{l(p)} = \text{Avg}[b_{l(p);i(p)}jl]$$

Location average of its industry premia $b_{l(p);i(p)}$

Weighting: plant employment w/in region (unweighted across regions) [Robustness: unweighted]

Main measure: TFPr (follow Foster et al. 2008) [Robustness: log value added per worker]

Raw Place E ects



Raw Place E ects

Raw Place E ects

Outline

Permutation Test

$$\underbrace{\mathrm{Var}(b^{S_l})}_{\text{Var of Est. Place E ects (Location Averages)}} = \underbrace{\mathrm{Var}(\ _l)}_{\text{Var of True Place E ects}} + \underbrace{\frac{1}{L}\sum_{l2L}\frac{_l(u)^2}{N^{S_l}}}_{\substack{>0 \text{ if } N<1^{\wedge} \quad (u)>0 \\ \text{Bias from Granularity:} \\ \text{Var of Sample Means}}} + \underbrace{2\,\mathrm{Cov}@_l;\ \frac{1}{N^{S_l}}\sum_{p2S_l} u_{pl}A}_{\substack{=0 \\ \text{Orthogonal by Construction}}}$$

Granularity-Bias-Only Benchmark: $F^{a}_{\ l}(a) = F^{a}(a) 8l\ 2\ L) \quad _l = \quad 8\,l\ 2\ L$

Implement via permutation test: randomly swap plants across MSAs within their industry

$$) \quad \underbrace{\mathrm{Var}(b^{S_l})}_{\text{Var of Est. Place E ects (Location Averages)}} = \underbrace{\mathrm{Var}(\ _l)}_{=0} + \underbrace{\frac{(a)^2}{N^{S_l}L}\sum_{l2L}}_{\substack{>0 \text{ if } N<1^{\wedge} \quad (a)>0 \\ \text{Bias from Granularity:} \\ \text{Var of Sample Means}}} + \underbrace{2\,\mathrm{Cov}@;\ \frac{1}{N^{S_l}}\sum_{p2S_l} a_{pl}A}_{\substack{=0 \\ \text{Orthogonal by Construction}}}$$

1,000 random US economies sampling distribution (in the dartboard spirit of Ellison & Glaeser 1997

Permutation Test: 1,000 Random Reallocations of Plants

Permutation Test: 1,000 Random Reallocations of Plants

Taking Stock

Taking Stock

Outline

Bias Correction of Variance: Split Samples

Split plants into two random and equally sized subsamples $s \in \{X, Y\}$ in each location l

Estimate two separate place averages for each location, $\bar{h}_l^X$; $\bar{h}_l^Y$

!!!! True place effect θ_l is common to both subsamples (by definition!)

Covariance of averages b/w subsamples is an unbiased estimator of variance of true place effects

$$\text{Cov}\left(\bar{h}_l^X ; \bar{h}_l^Y\right) = \text{Cov}\left(\theta_l + u_l^{X_l} ; \theta_l + u_l^{Y_l}\right)$$

$$= \underbrace{\text{Var}(\theta_l)}_{\text{Var of True Place Effects}} + \underbrace{\text{Cov}\left(\theta_l ; u_l^{X_l}\right)}_{=0} + \underbrace{\text{Cov}\left(\theta_l ; u_l^{Y_l}\right)}_{=0} + \underbrace{\text{Cov}\left(u_l^{X_l} ; u_l^{Y_l}\right)}_{=0}$$

(where $u_l^{S_l} = \frac{1}{N^{S_l}} \sum_{p \in S_l} u_{pl}$)

Bias Correction of Variance: Split Samples

Bias Correction of Variance: Split Samples

$= \frac{\text{Cov}(b^Y; b^X)}{\text{Var}(b^X)} = \frac{\text{Var}(\ )}{\text{Var}(b^X)} =$ i.e. the share of variance of true place e ects buried in raw variance

Bias Correction of Variance: Split Samples

$= \frac{\text{Cov}(b^Y; b^X)}{\text{Var}(b^X)} = \frac{\text{Var}(\ )}{\text{Var}(b^X)} =$ i.e. the share of variance of true place e ects buried in raw variance

Bias Correction of Variance: Split Samples

$= \frac{\text{Cov}(b^Y; b^X)}{\text{Var}(b^X)} = \frac{\text{Var}(\ )}{\text{Var}(b^X)} =$ i.e. the share of variance of true place e ects buried in raw variance

Bias Correction of Variance: Split Samples

$$= \frac{\mathrm{Cov}(b^Y; b^X)}{\mathrm{Var}(b^X)} = \frac{\mathrm{Var}(\ )}{\mathrm{Var}(b^X)} =$$ i.e. the share of variance of true place e ects buried in raw variance

See paper: Sample split leads to higher raw variance by doubling granularity bias on x-axis, so is lower than the variance ratios on next slide using the full sample $\mathrm{Var}(b^{X[Y})$ to compute raw variance .

Bias Correction of Variance: Split Samples

Bias Correction of Variance: Split Samples

66% of raw variance in location e ects is spurious { "granularity bias"

, At most 1/3 re ects variance of true place e ects

Outline

De nitions & statistical basics

Raw variance

Permutation test: granularity-bias-only benchmark

Bias correction: split-sample method

Tracing the sources of granularity bias

Extension I: within-country dispersion in 15 European countries

Extension II: new plants' place e ects

Sources of Granularity Bias

$$\underbrace{\mathrm{Var}(b^{S_l})}_{\text{Var of Est. Place E ects (Location Averages)}} = \underbrace{\mathrm{Var}(\ _l)}_{\text{Var of True Place E ects}} + \underbrace{\frac{1}{L}\sum_{l2L}\frac{_l(u)^2}{N^{S_l}}}_{\text{Bias from Granularity: Var of Sample Means}} + \underbrace{2\,\mathrm{Cov}@_l;\ \frac{1}{N^{S_l}}\sum_{p2S_l} u_{pl}\,A}_{\text{Orthogonal by Construction}}$$

> 0 if $N < 1^{\wedge}$ $(u) > 0$

$=0$

Three sources

- Idiosyncratic plant heterogeneity
- Finite samples of places within the place
- Weighting: large plants

for exposition, equation is unweighted but baseline implementatio weighted

In the paper, we dissect each source in dedicated checks

Cutting Away Granular Cells

Cutting Away Granular Cells

Cutting Away Granular Cells

Outline

De nitions & statistical basics

Raw variance

Permutation test: granularity-bias-only benchmark

Bias correction: split-sample method

Tracing the sources of granularity bias

Extension I: within-country dispersion in 15 European countries

Extension II: new plants' place e ects

New Plants: Even Higher Raw Variance

New Plants: Even Higher Raw Variance

New Plants: Even Higher Bias

Bias corrected variances are very similar between new (0.015, 0.005) and old (0.008, 0.009).

New Plants: Covariance With Old Plants

New Plants: Covariance With Old Plants

New Plants: Covariance With Old Plants

Outline

De nitions & statistical basics

Raw variance

Permutation test: granularity-bias-only benchmark

Bias correction: split-sample method

Tracing the sources of granularity bias

Extension I: within-country dispersion in 15 European countries

Extension II: new plants' place e ects { 5 years and younger

15 European Countries: Location E ects

Measuring $\mathrm{Var}(\underbrace{\text{Location-Specific } E[\text{Plant-Level Productivity}]}_{\text{"True Place Effects"}})$

Places do differ significantly in plant productivity.

Large raw variance, but:

- At least 3/4 is spurious (granularity bias: idiosyncratic plant-level dispersion in productivity).
- 1/4 due to true place effects.

Removing most granular cells reduces raw variance but also \true" variance.

Patterns extend to 15 European countries.

Bias larger for new plants.

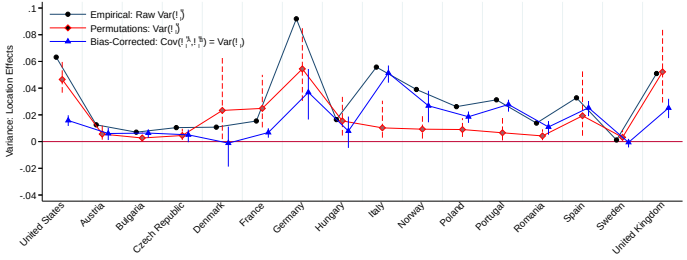
Place effects for new plants somewhat distinct from those of old plants.

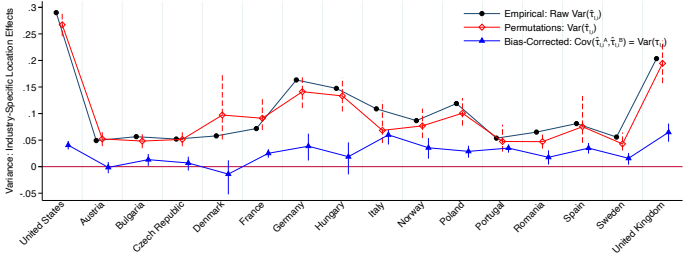
Appendix Slides

	Main	unw'd	wnwin.	2.5% win.	6-d	10	New&Old	New	Old
Panel A: Variance of Place E ects									
Raw var.	0.026	0.005	0.029	0.022	0.014	0.021	0.032	0.084	0
Perm. mean	0.017	0.003	0.019	0.014	0.014	0.014	0.025	0.062	0
Perm. sd	0.003	0.000	0.004	0.002	0.003	0.002	0.004	0.010	0
p-val	0.009	0.003	0.016	0.004	0.424	0.016	0.055	0.021	0
Cov. mean	0.006	0.002	0.007	0.006	0.002	0.008	0.008	0.015	0
Cov. UB	0.009	0.003	0.010	0.008	0.003	0.010	0.013	0.024	0
Cov. LB	0.004	0.001	0.004	0.003	0.000	0.005	0.004	0.003	0
Panel B: Variance of Place-Industry E ects									
Raw var.	0.110	0.049	0.124	0.094	0.072	0.044	0.064	0.152	0
Perm. mean	0.096	0.042	0.110	0.083	0.071	0.035	0.050	0.127	0
Perm. sd.	0.005	0.001	0.007	0.004	0.005	0.003	0.005	0.011	0
p-val	0.015	0.001	0.037	0.005	0.351	0.004	0.007	0.018	0
Cov. mean	0.007	0.005	0.008	0.006	0.002	0.008	0.009	0.005	0
Cov. UB	0.009	0.006	0.010	0.008	0.003	0.009	0.011	0.008	0
Cov. LB	0.006	0.004	0.007	0.005	0.001	0.006	0.007	0.002	0
N MSAs	380	380	380	380	380	250	300	300	3
N ind-MSAs	11500	11500	11500	11500	18000	2800	2800	2800	2
N	120000	120000	120000	120000	105000	86000	78000	14000	6

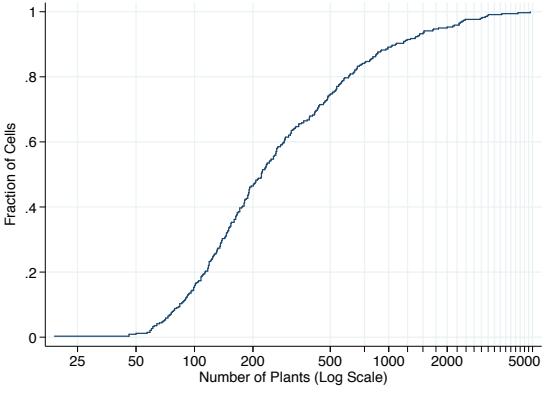
	Main	unw'd	wnwin.	2.5% win.	6-d	10	New&Old	New	Old
Panel A: Variance of Place E ects									
Raw var.	0.063	0.018	0.069	0.053	0.073	0.052	0.085	0.151	0
Perm. mean	0.046	0.008	0.053	0.039	0.051	0.038	0.068	0.165	0
Perm. sd	0.006	0.001	0.008	0.005	0.007	0.006	0.009	0.019	0
p-val	0.008	0.000	0.041	0.009	0.003	0.020	0.047	0.754	0
Cov. mean	0.019	0.010	0.020	0.015	0.027	0.018	0.019	0.013	0
Cov. UB	0.024	0.012	0.026	0.020	0.032	0.024	0.027	0.027	0
Cov. LB	0.013	0.008	0.013	0.010	0.021	0.012	0.011	-0.002	0
Panel B: Variance of Place-Industry E ects									
Raw var.	0.290	0.132	0.328	0.244	0.288	0.113	0.163	0.322	0
Perm. mean	0.268	0.110	0.310	0.226	0.259	0.092	0.138	0.338	0
Perm. sd	0.011	0.003	0.016	0.008	0.011	0.007	0.011	0.022	0
p-val	0.028	0.000	0.131	0.020	0.007	0.003	0.028	0.754	0
Cov. mean	0.031	0.020	0.037	0.027	0.021	0.023	0.030	0.029	0
Cov. UB	0.035	0.022	0.041	0.030	0.023	0.027	0.035	0.036	0
Cov. LB	0.028	0.017	0.033	0.024	0.019	0.019	0.024	0.023	0
N MSAs	380	380	380	380	380	250	300	300	3
N ind-MSAs	11500	11500	11500	11500	18000	2800	2800	2800	
N	120000	120000	120000	120000	105000	86000	78000	14000	

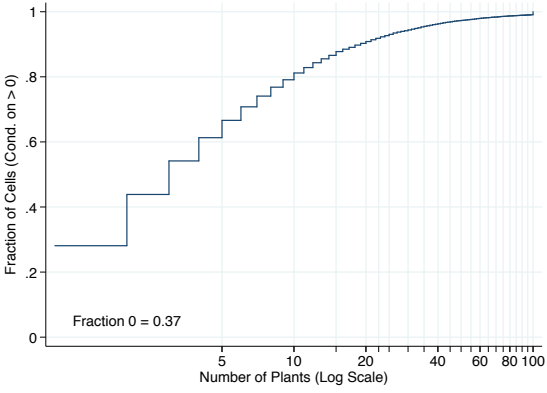
Log Value Added Per Worker





Cell Counts


1
.8
.6
.4
.2
0
Fraction of Cells
25
50
100
200
500
1000
2000
5000
Number of Plants (Log Scale)


1
.8
.6
.4
.2
0
Fraction of Cells (Cond. on > 0)
Fraction 0 = 0.37
5
10
20
40
60
80
100
Number of Plants (Log Scale)