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 differ 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

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

- **Key findings:**
  - 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

  $\Leftrightarrow$ Much smaller true place effects: at most 1/4 - 1/3 of raw cross-MSA variance reflects true place effects

- **Applications/extensions**:
  - Robustness...
  - New plants - 60% pass-through of true place effects.
  - Extend to w/in regional dispersion in 15 European countries

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

Definitions & Basic Statistics

- Plant p in location $l \in L$ has productivity (log TFP) a_{pl} .
- $a \sim F_l^a(a)$ – DGP is l-specific.
  - Statistical def, agnostic to sources of l-dependence: sorting, agglomeration effects, mismeasurement,...
- True place effect:

$$\begin{aligned} \tau_l &= \mathsf{E}\left[a_{pl} \middle| l\right] \\ &= \int a dF_l^s(a) \\ \Rightarrow a_{pl} &= \tau_l + u_{pl} \end{aligned}$$

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

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

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

Raw Variance of Place Averages

$$\overbrace{\text{Var}(\widehat{\tau}_l^{S_l})}^{\substack{\text{Var of Est. Place Effects} \\ \text{(Location Averages)}}} = \text{Var}\left(\frac{1}{N^{S_l}} \sum_{p \in S_l} a_{pl}\right)$$

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

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

Pitfalls: Granularity Bias

$$\overbrace{\text{Var}(\widehat{\tau}_l^{S_l})}^{\substack{\text{Var of Est. Place Effects}\\ \text{(Location Averages)}}} = \text{Var}\left(\frac{1}{N^{S_l}}\sum_{p\in S_l} a_{pl}\right)$$

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

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

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

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:

$$\widehat{\tau}_{l(p),i(p)} = \text{Avg}[a_{pl}|i,l] - \bar{a}_i$$

  - Demeaned $\Rightarrow$ "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]

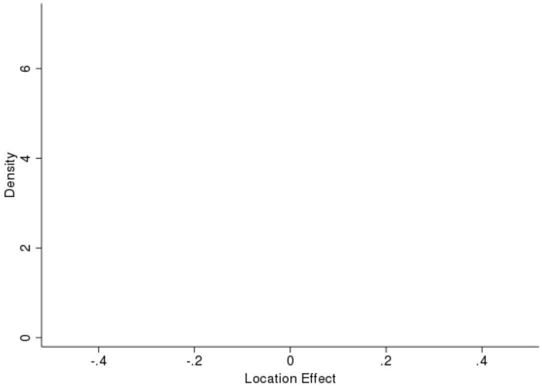
- Overall location effect:

$$\widehat{\xi}_{l(p)} = \text{Avg}[\widehat{\tau}_{l(p),i(p)}|l]$$

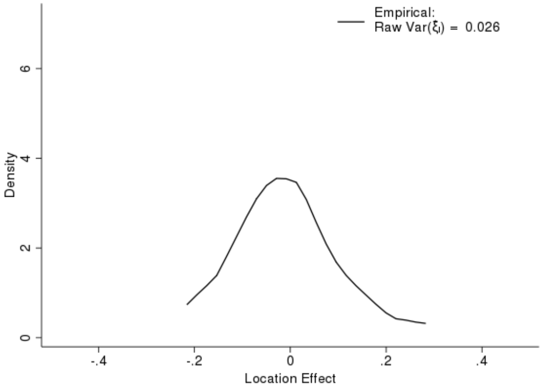
  ∼ Location average of its industry premia $\widehat{\tau}_{l(p),i(p)}$

- 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]

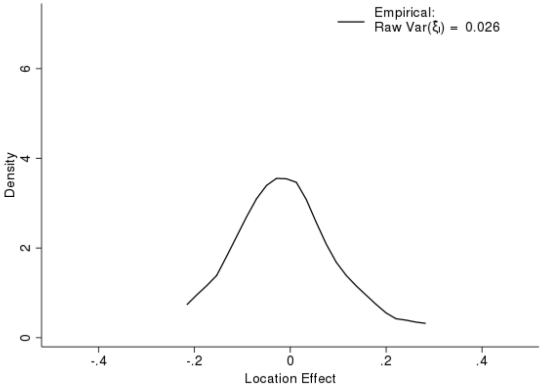
Raw Place Effects

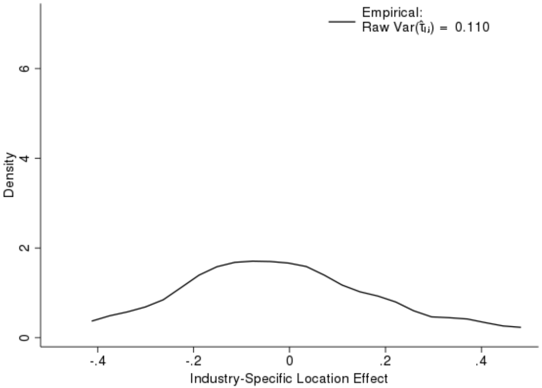


Raw Place Effects



Raw Place Effects





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

Permutation Test

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

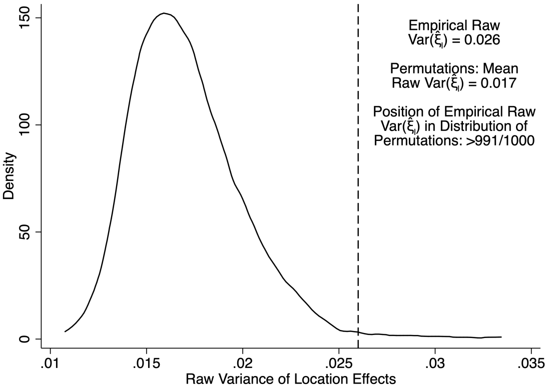
Granularity-Bias-Only Benchmark: $F_l^a(a) = F^a(a) \forall l \in L \Rightarrow \tau_l = \tau \ \ \forall\, l \in L$

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

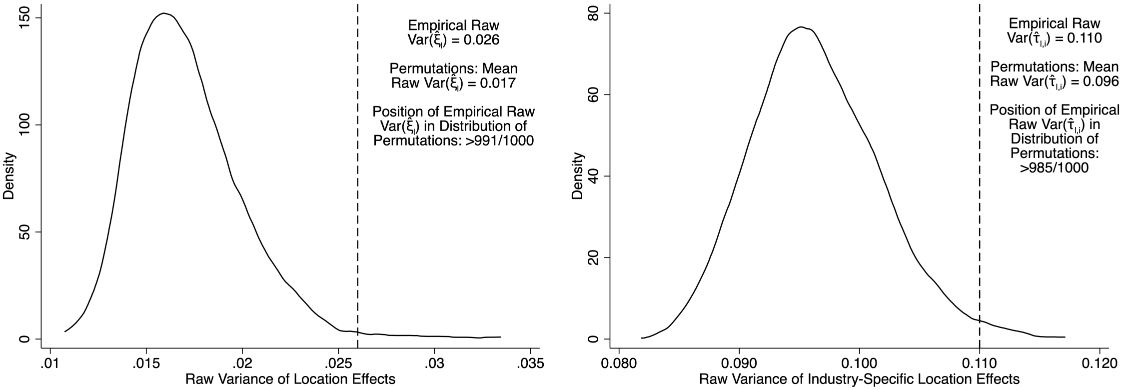
$$\Rightarrow \underbrace{\text{Var}(\widehat{\tau}_l^{S_l})}_{\substack{\text{Var of Est. Place Effects}\\ \text{(Location Averages)}}} = \underbrace{\text{Var}(\tau)}_{=0} + \underbrace{\sigma(a)^2 \cdot \sum_{l\in L}\frac{1}{N^{S_l}L}}_{\substack{>0 \text{ if } N<\infty \wedge \sigma(a)>0\\ \text{Bias from Granularity:}\\ \text{Var of Sample Means}}} + \underbrace{2\,\text{Cov}\left(\tau, \frac{1}{N^{S_l}}\sum_{p\in S_l} a_{pl}\right)}_{\substack{=0\\ \text{Orthogonal by Construction}}}$$

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

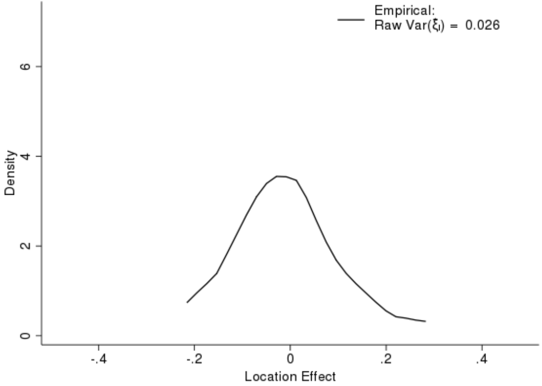
Permutation Test: 1,000 Random Reallocations of Plants

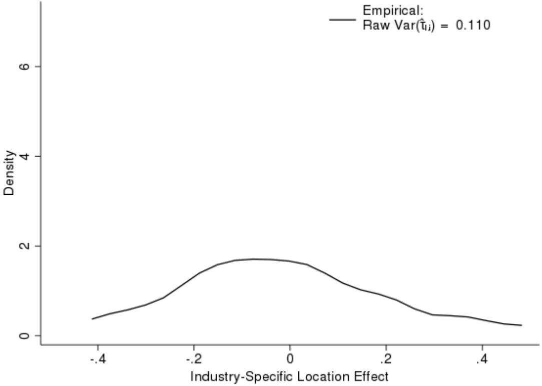


Permutation Test: 1,000 Random Reallocations of Plants

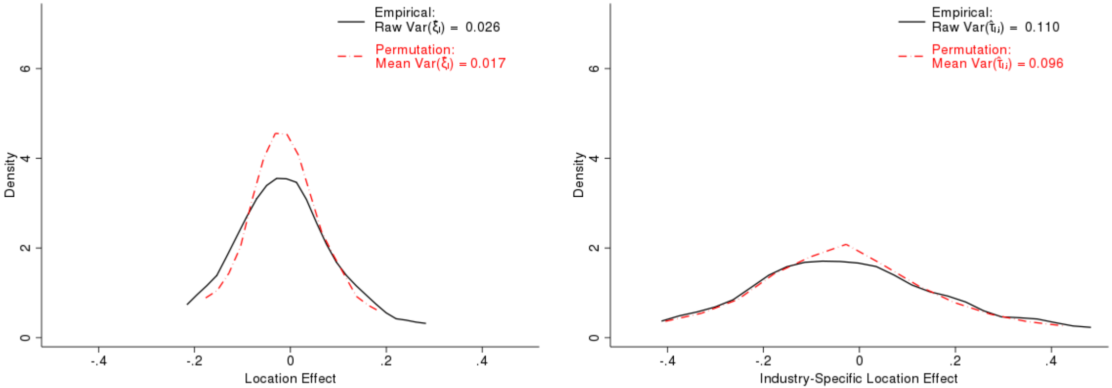


Taking Stock





Taking Stock



Outline

- 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

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 l, $\widehat{\tau_l}^X, \widehat{\tau_l}^Y$

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

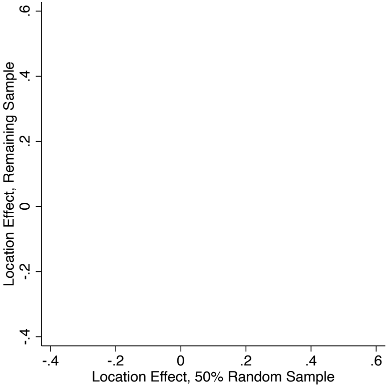
- Covariance of averages b/w subsamples is an unbiased estimator of variance of true place effects

$$\text{Cov}\left(\widehat{\tau}_l^X, \widehat{\tau}_l^Y,\right) = \text{Cov}\left(\tau_l + \bar{u}_l^{X_l}, \tau_l + \bar{u}_l^{Y_l}\right)$$

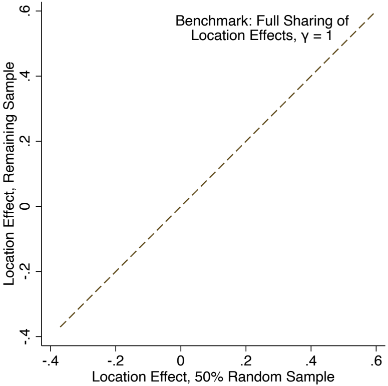
$$= \underbrace{\text{Var}\left(\tau_l\right)}_{\text{Var of True Place Effects}} + \underbrace{\text{Cov}\left(\tau_l, \bar{u}_l^{X_l}\right)}_{=0} + \underbrace{\text{Cov}\left(\tau_l, \bar{u}_l^{Y_l}\right)}_{=0} + \underbrace{\text{Cov}\left(\bar{u}_l^{X_l}, \bar{u}_l^{Y_l}\right)}_{=0}$$

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

Bias Correction of Variance: Split Samples

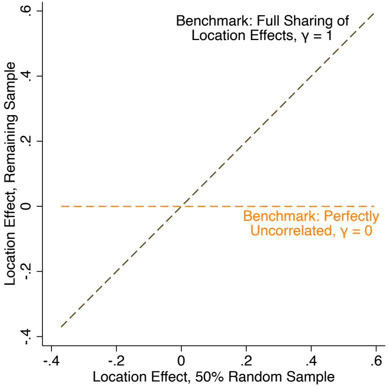


Bias Correction of Variance: Split Samples



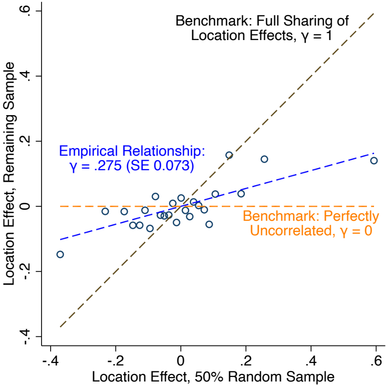
$$\gamma = \frac{\text{Cov}(\widehat{\tau}^Y, \widehat{\tau}^X)}{\text{Var}(\widehat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\widehat{\tau}^X)} = \text{ i.e. the share of variance of true place effects buried in raw variance}$$

Bias Correction of Variance: Split Samples



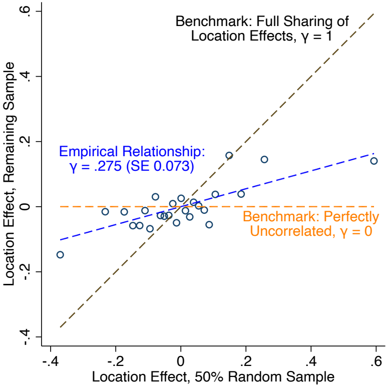
$$\gamma = \frac{\text{Cov}(\widehat{\tau}^Y, \widehat{\tau}^X)}{\text{Var}(\widehat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\widehat{\tau}^X)} =$$ i.e. the share of variance of true place effects buried in raw variance

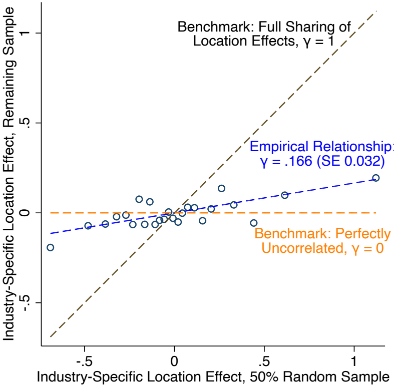
Bias Correction of Variance: Split Samples



$$\gamma = \frac{\text{Cov}(\widehat{\tau}^Y, \widehat{\tau}^X)}{\text{Var}(\widehat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\widehat{\tau}^X)} =$$ i.e. the share of variance of true place effects buried in raw variance

Bias Correction of Variance: Split Samples

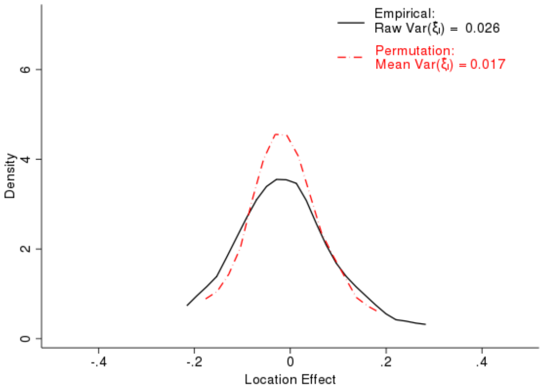


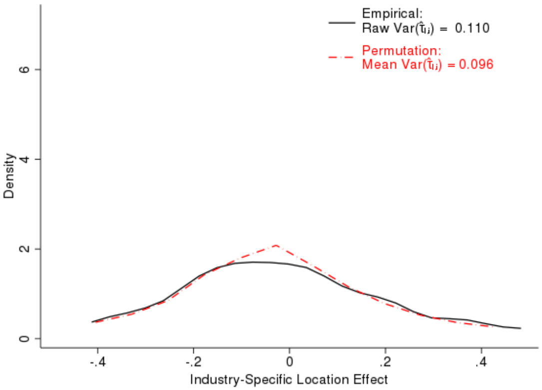


$$\gamma = \frac{\text{Cov}(\widehat{\tau}^Y, \widehat{\tau}^X)}{\text{Var}(\widehat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\widehat{\tau}^X)} = $$ i.e. the share of variance of true place effects buried in raw variance

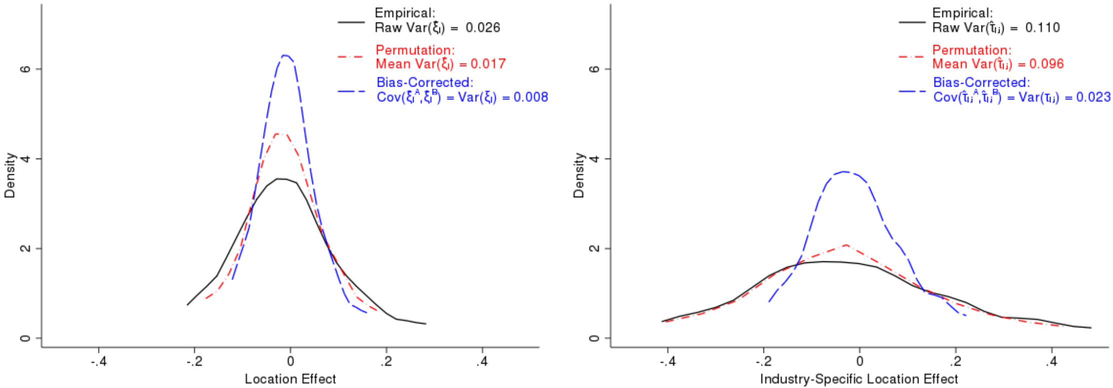
See paper: Sample split leads to higher raw variance by doubling granularity bias on x-axis, so γ is lower than the variance ratios on next slide using the full sample $\text{Var}(\widehat{\tau}^{X \cup Y})$ to compute raw variance .

Bias Correction of Variance: Split Samples





Bias Correction of Variance: Split Samples



- 66% of raw variance in location effects is spurious – "granularity bias"

⇔ At most 1/3 reflects variance of true place effects

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: within-country dispersion in 15 European countries
- Extension II: new plants' place effects

Sources of Granularity Bias

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

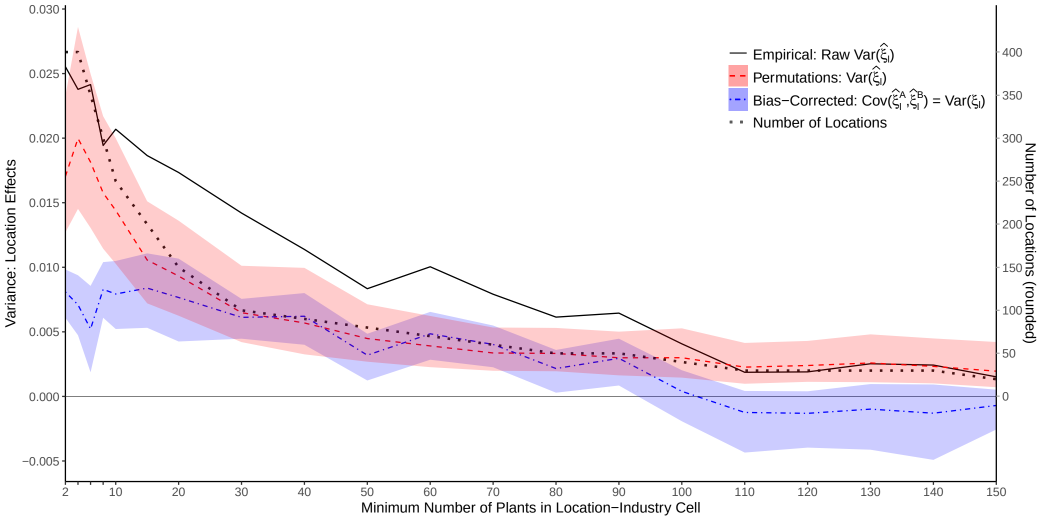
Three sources

- Idiosyncratic plant heterogeneity
- Finite samples of places within the place
- Weighting: large plants for exposition, equation is unweighted but baseline implementation is weighted

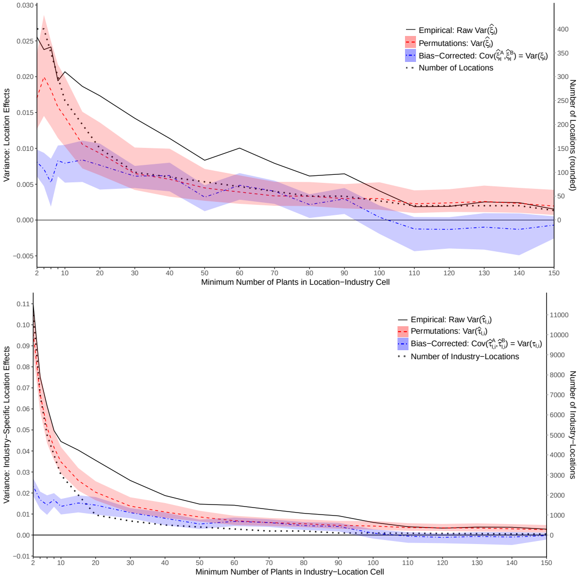
In the paper, we dissect each source in dedicated checks

Cutting Away Granular Cells

Cutting Away Granular Cells



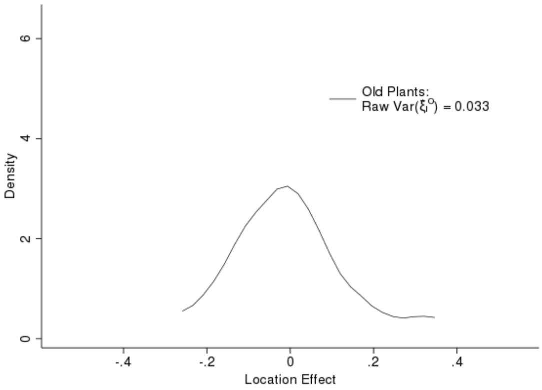
Cutting Away Granular Cells

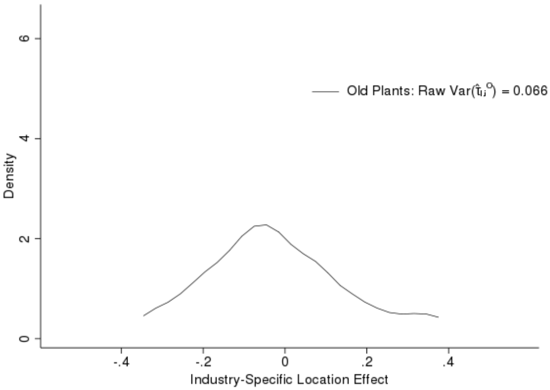


Outline

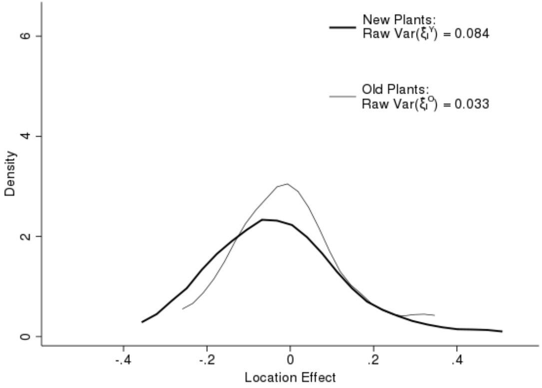
- Definitions & statistical basics
- Raw variance
- Permutation test: granularity-bias-only benchmark
- Bias correction: split-sample method
- Tracing the sources of granularity bias
- Extension I: within-country dispersion in 15 European countries
- Extension II: new plants' place effects

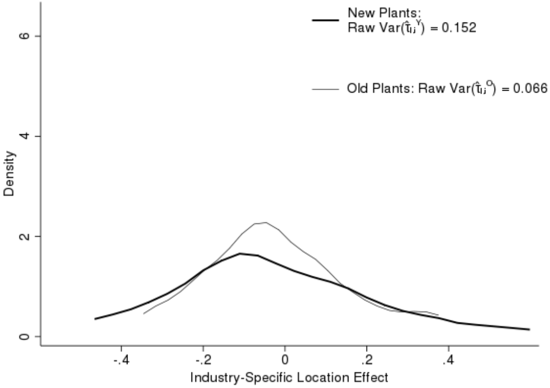
New Plants: Even Higher Raw Variance



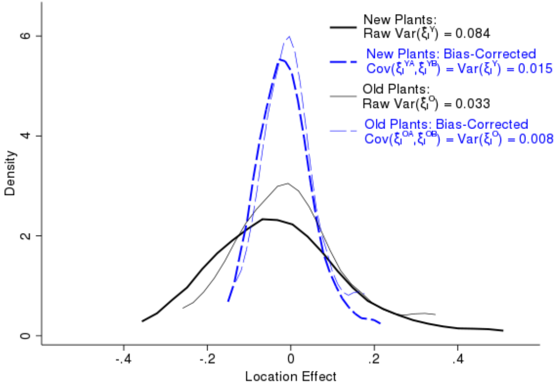


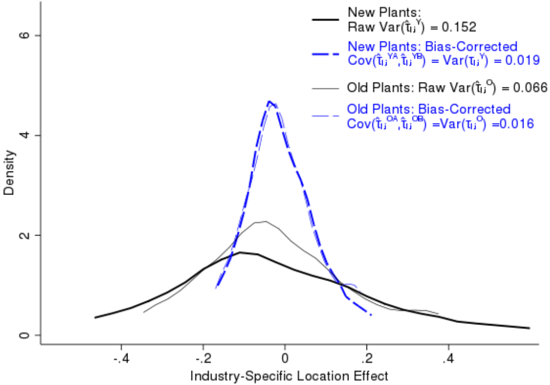
New Plants: Even Higher Raw Variance





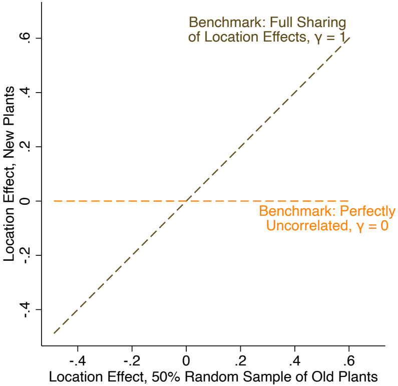
New Plants: Even Higher Bias

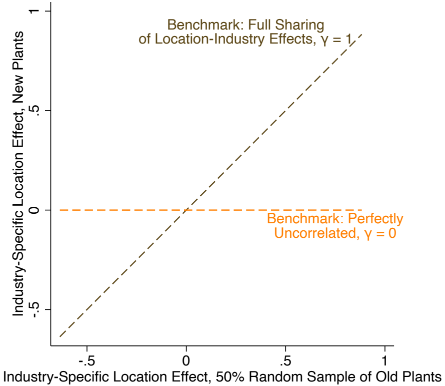




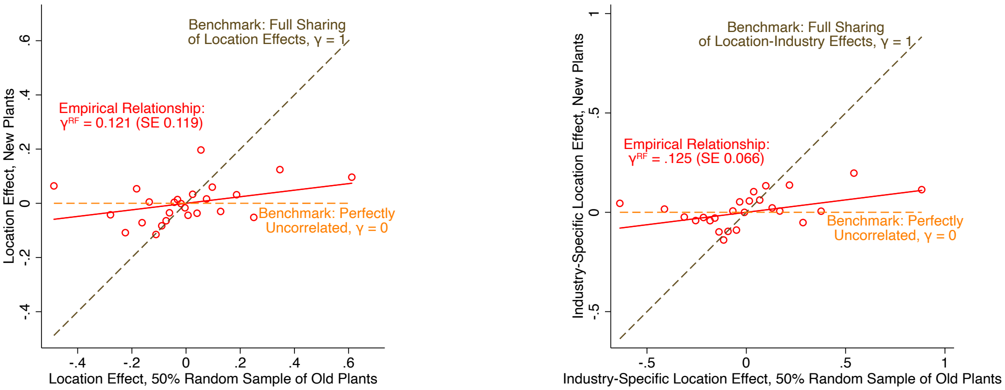
Bias corrected variances are very similar between new (0.015, 0.005) and old (0.008, 0.009).

New Plants: Covariance With Old Plants

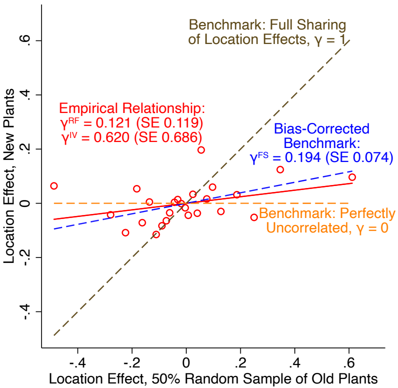


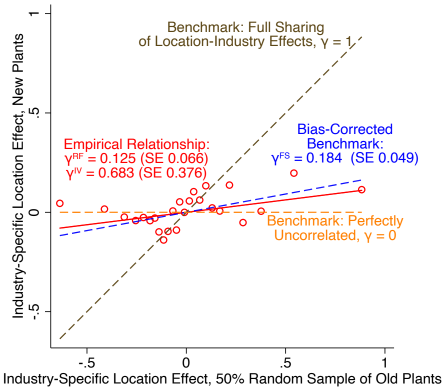


New Plants: Covariance With Old Plants



New Plants: Covariance With Old Plants

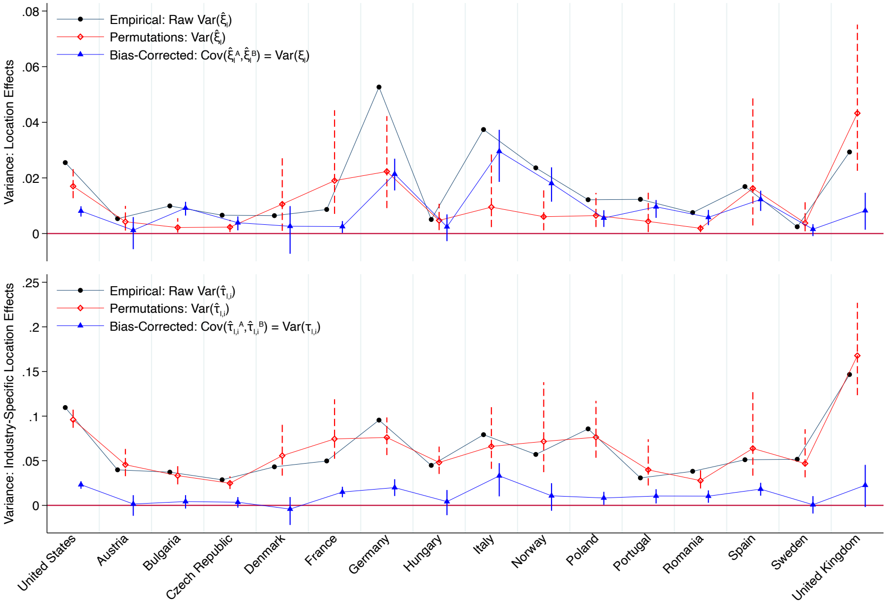




Outline

- Definitions & 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 effects – 5 years and younger

15 European Countries: Location Effects



Measuring Var(**Location-Specific** E[**Plant-Level Productivity**])

$\underbrace{\qquad\qquad}_{\text{"True Place Effects"}}$

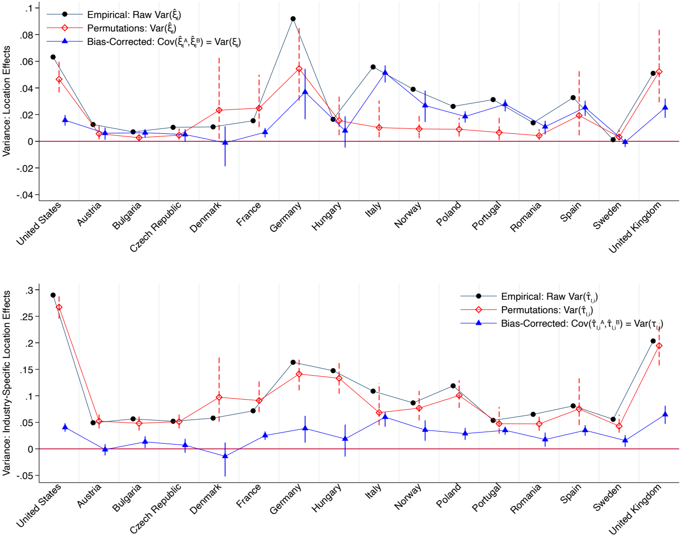
- Places do differ significantly in plant productivity.
- Large raw variance, but:
  - At least 3/4 is spurious (granularity bias: idiosyncratic plant-level dispersion in productivity).
  - $\Leftrightarrow$ 1/4 due to true place effects.
- Removing most granular cells reduces raw variance but also "true" 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

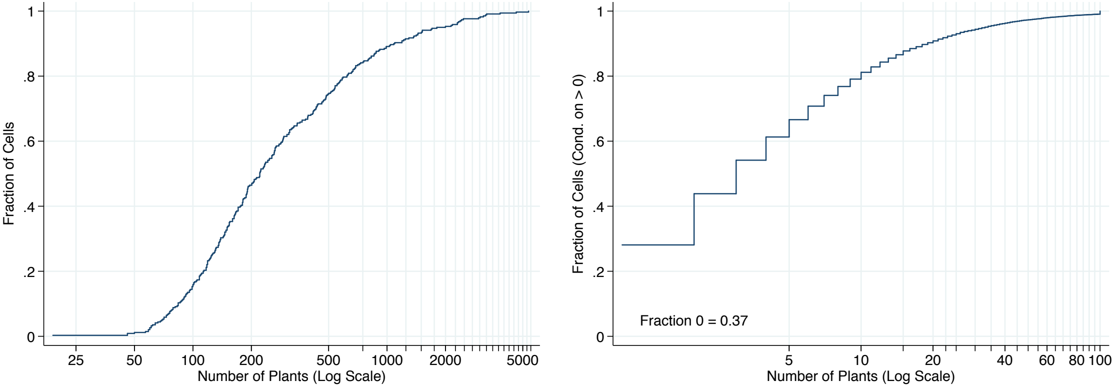
	Main	unw'd	wnwin.	2.5% win.	6-d	≥ 10	New&Old	New	Old
Panel A: Variance of Place Effects									
Raw var.	0.026	0.005	0.029	0.022	0.014	0.021	0.032	0.084	0.033
Perm. mean	0.017	0.003	0.019	0.014	0.014	0.014	0.025	0.062	0.032
Perm. sd	0.003	0.000	0.004	0.002	0.003	0.002	0.004	0.010	0.006
p-val	0.009	0.003	0.016	0.004	0.424	0.016	0.055	0.021	0.418
Cov. mean	0.006	0.002	0.007	0.006	0.002	0.008	0.008	0.015	0.008
Cov. UB	0.009	0.003	0.010	0.008	0.003	0.010	0.013	0.024	0.012
Cov. LB	0.004	0.001	0.004	0.003	0.000	0.005	0.004	0.003	0.004
Panel B: Variance of Place-Industry Effects									
Raw var.	0.110	0.049	0.124	0.094	0.072	0.044	0.064	0.152	0.066
Perm. mean	0.096	0.042	0.110	0.083	0.071	0.035	0.050	0.127	0.061
Perm. sd.	0.005	0.001	0.007	0.004	0.005	0.003	0.005	0.011	0.007
p-val	0.015	0.001	0.037	0.005	0.351	0.004	0.007	0.018	0.233
Cov. mean	0.007	0.005	0.008	0.006	0.002	0.008	0.009	0.005	0.009
Cov. UB	0.009	0.006	0.010	0.008	0.003	0.009	0.011	0.008	0.011
Cov. LB	0.006	0.004	0.007	0.005	0.001	0.006	0.007	0.002	0.007
N MSAs	380	380	380	380	380	250	300	300	300
N ind-MSAs	11500	11500	11500	11500	18000	2800	2800	2800	2800
N	120000	120000	120000	120000	105000	86000	78000	14000	64000

	Main	unw'd	wnwin.	2.5% win.	6-d	≥ 10	New&Old	New	Old
				Panel A: Variance of Place Effects					
Raw var.	0.063	0.018	0.069	0.053	0.073	0.052	0.085	0.151	0.101
Perm. mean	0.046	0.008	0.053	0.039	0.051	0.038	0.068	0.165	0.086
Perm. sd	0.006	0.001	0.008	0.005	0.007	0.006	0.009	0.019	0.012
p-val	0.008	0.000	0.041	0.009	0.003	0.020	0.047	0.754	0.112
Cov. mean	0.019	0.010	0.020	0.015	0.027	0.018	0.019	0.013	0.018
Cov. UB	0.024	0.012	0.026	0.020	0.032	0.024	0.027	0.027	0.028
Cov. LB	0.013	0.008	0.013	0.010	0.021	0.012	0.011	-0.002	0.007
				Panel B: Variance of Place-Industry Effects					
Raw var.	0.290	0.132	0.328	0.244	0.288	0.113	0.163	0.322	0.182
Perm. mean	0.268	0.110	0.310	0.226	0.259	0.092	0.138	0.338	0.164
Perm. sd	0.011	0.003	0.016	0.008	0.011	0.007	0.011	0.022	0.014
p-val	0.028	0.000	0.131	0.020	0.007	0.003	0.028	0.754	0.091
Cov. mean	0.031	0.020	0.037	0.027	0.021	0.023	0.030	0.029	0.018
Cov. UB	0.035	0.022	0.041	0.030	0.023	0.027	0.035	0.036	0.023
Cov. LB	0.028	0.017	0.033	0.024	0.019	0.019	0.024	0.023	0.013
N MSAs	380	380	380	380	380	250	300	300	300
N ind-MSAs	11500	11500	11500	11500	18000	2800	2800	2800	2800
N	120000	120000	120000	120000	105000	86000	78000	14000	64000

Log Value Added Per Worker

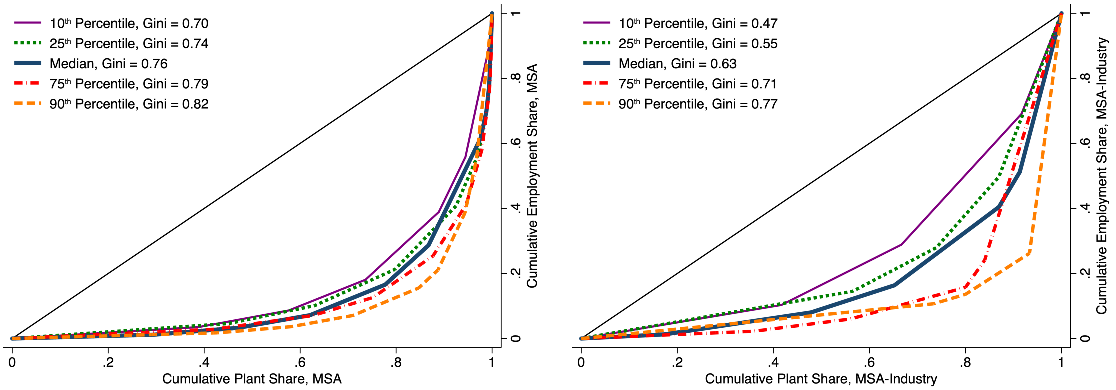


Cell Counts



Fraction of Cells
1
.8
.6
.4
.2
0
25
50
100
200
500
1000
2000
5000
Number of Plants (Log Scale)
Fraction of Cells (Cond. on > 0)
1
.8
.6
.4
.2
0
Fraction 0 = 0.37
5
10
20
40
60
80
100
Number of Plants (Log Scale)

Plant Size



10th Percentile, Gini = 0.70
25th Percentile, Gini = 0.74
Median, Gini = 0.76
75th Percentile, Gini = 0.79
90th Percentile, Gini = 0.82
Cumulative Employment Share, MSA
Cumulative Plant Share, MSA
0
.2
.4
.6
.8
1
10th Percentile, Gini = 0.47
25th Percentile, Gini = 0.55
Median, Gini = 0.63
75th Percentile, Gini = 0.71
90th Percentile, Gini = 0.77
Cumulative Employment Share, MSA-Industry
Cumulative Plant Share, MSA-Industry