Productivity, Place, and Plants: Revisiting the Measurement

Preliminary: update pending due to Covid-related Census RDC closures.

Benjamin Schoefer
UC Berkeley

Oren Ziv
Michigan State University

NBER Summer Institute 2020
Urban Economics

Thanks a lot to C. Luke Watson (MSU PhD '21) for fantastic research assistance!

Measuring Spatial Di erences in Productivity

Q: How much do plants differ in their productivity across places for reasons that are systematically related to their current location?

Goal: isolate cross-regional variance in true place effects , strip out bias from idiosyncratic plant-level heterogeneity ("granularity bias")

US MSAs; Census of Manufactures

Extension I: new plants

Extension II: 15 European countries (Bureau van Dijk)

Key nding:

2/3 of raw variance is spurious { granularity bias

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

Outline

De nitions & Basic Statistics

Plant p in location l 2 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:

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

Measured average productivity of N^{S_l} plants p 2 S_l:

$$\widehat{\tau}_l^{S_l} = \frac{1}{N^{S_l}} \sum_{p2S_l} a_{pl}.$$

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

Raw Variance of Place Averages

$$\overbrace{\text{Var}(\widehat{\tau}_l^{S_l})}^{\substack{\text{Var of Est. Place Effects} \\ \text{(Location Averages)}}} = \text{Var}\left(\frac{1}{N^{S_l}}\sum_{p2S_l} a_{pl}\right)$$

$$= \text{Var}\left(\frac{1}{N^{S_l}}\sum_{p2S_l} [\tau_l + u_{pl}]\right)$$

$$= \text{Var}\left(\tau_l + \frac{1}{N^{S_l}}\sum_{p2S_l} u_{pl}\right)$$

Pitfalls: Granularity Bias

Var of Est. Place Effects (Location Averages)

$$\overbrace{\text{Var}(\widehat{\tau}_l^{S_l})}$$

$$= \text{Var}\left(\frac{1}{N^{S_l}}\sum_{p2S_l} a_{pl}\right)$$

$$= \text{Var}\left(\frac{1}{N^{S_l}}\sum_{p2S_l} [\tau_l + u_{pl}]\right)$$

$$= \text{Var}\left(\tau_l + \frac{1}{N^{S_l}}\sum_{p2S_l} u_{pl}\right)$$

$$= \underbrace{\text{Var}(\tau_l)}_{\text{Var of True Place Effects}} + \underbrace{\frac{1}{L}\sum_{l2L}\frac{\sigma_l(u)^2}{N^{S_l}}}_{\substack{>0 \text{ if } N<1 \text{ ^}\sigma(u)>0 \\ \text{Bias from Granularity:} \\ \text{Var of Sample Means}}} + \underbrace{2\,\text{Cov}\left(\tau_l, \frac{1}{N^{S_l}}\sum_{p2S_l} u_{pl}\right)}_{\substack{=0 \\ \text{Orthogonal by Construction}}}$$

Outline

Definitions & statistical basics

Raw variance

Permutation test: granularity-bias-only benchmark

Bias correction: split-sample method

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:

$$\widehat{\tau}_{l(p),i(p)} = \text{Avg}[a_{pl} | i, l] \quad a_i$$

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

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

5 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 $\widehat{\ }_{l(p),i(p)}$

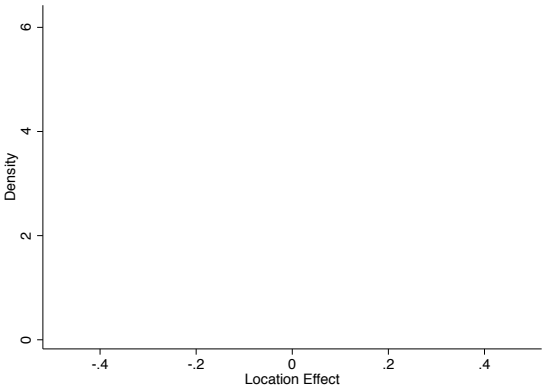
Weighting: plant employment [Robustness: unweighted]

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

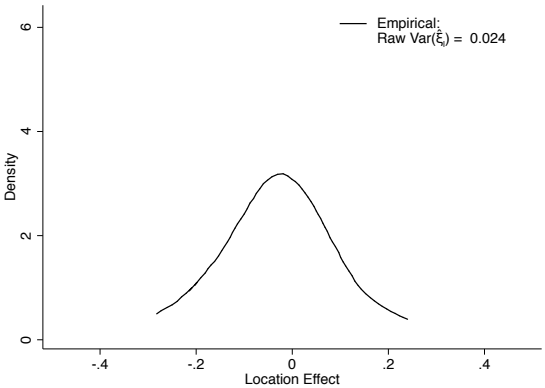
Estimate via fixed effects regressions

Currently: pool 5 Census waves demeaned by year. Post-Covid reeopening of Census RDCs, only 2012 wave.

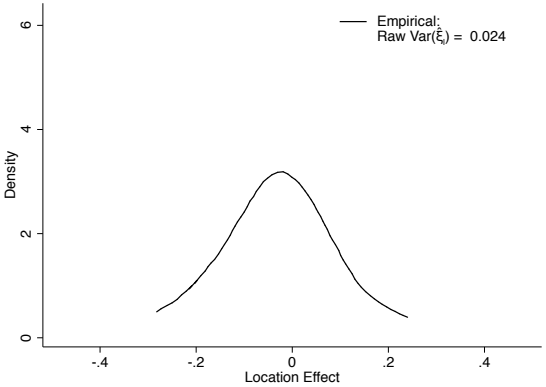
Raw Place E ects

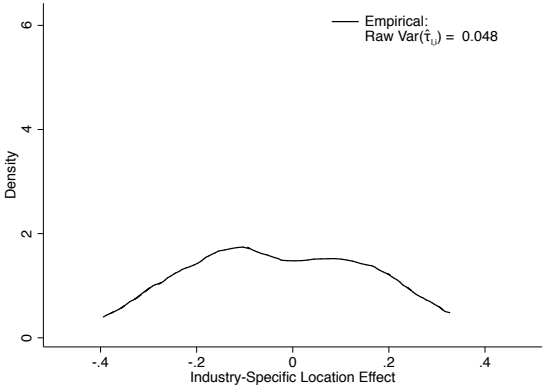


Raw Place E ects



Raw Place E ects





Outline

Definitions & statistical basics

Raw variance

Permutation test: granularity-bias-only benchmark

Bias correction: split-sample method

Extension I: new plants' place effects

Extension II: within-country dispersion in 15 European countries

Permutation Test

$$\underbrace{\text{Var}(\widehat{\tau}_l^{S_l})}_{\substack{\text{Var of Est. Place Effects} \\ \text{(Location Averages)}}} = \underbrace{\text{Var}(\tau_l)}_{\text{Var of True Place Effects}} + \underbrace{\frac{1}{L}\sum_{l2L}\frac{\sigma_l(u)^2}{N^{S_l}}}_{\substack{>0 \text{ if } N<1 \ \wedge \sigma(u)>0 \\ \text{Bias from Granularity:} \\ \text{Var of Sample Means}}} + \underbrace{2\,\text{Cov}\left(\tau_l, \frac{1}{N^{S_l}}\sum_{p2S_l} u_{pl}\right)}_{\substack{=0 \\ \text{Orthogonal by Construction}}}$$

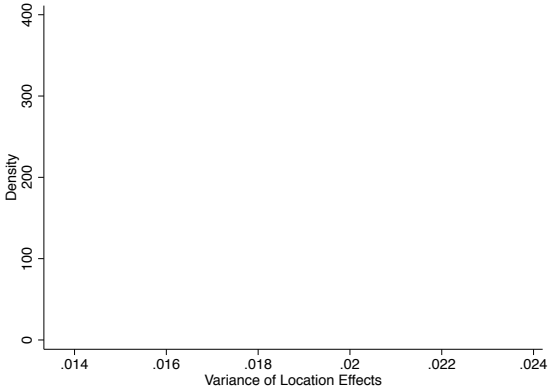
Granularity-Bias-Only Benchmark: $F_l^a(a) = F^a(a) 8l\ 2\ L\ ) \ \tau_l = \tau\ \ 8\ l\ 2\ L$

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

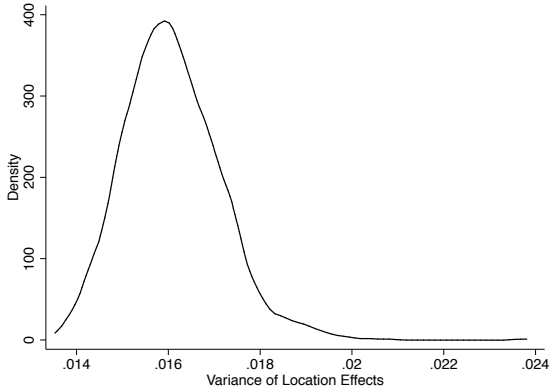
$$) \quad \underbrace{\text{Var}(\widehat{\tau}_l^{S_l})}_{\substack{\text{Var of Est. Place Effects} \\ \text{(Location Averages)}}} = \underbrace{\text{Var}(\tau)}_{=0} + \underbrace{\sigma(a)^2 \sum_{l2L}\frac{1}{N^{S_l}L}}_{\substack{>0 \text{ if } N<1 \ \wedge \sigma(a)>0 \\ \text{Bias from Granularity:} \\ \text{Var of Sample Means}}} + \underbrace{2\,\text{Cov}\left(\tau, \frac{1}{N^{S_l}}\sum_{p2S_l} a_{pl}\right)}_{\substack{=0 \\ \text{Orthogonal by Construction}}}$$

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

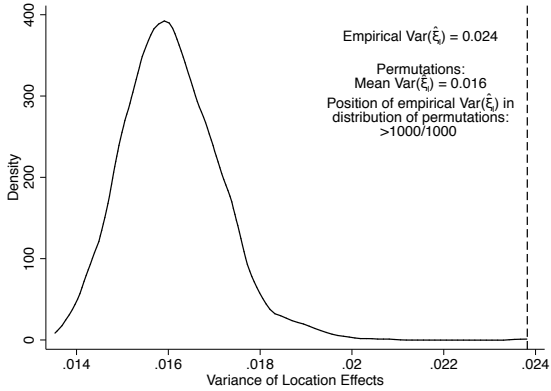
Permutation Test: 1,000 Random Reallocations of Plants



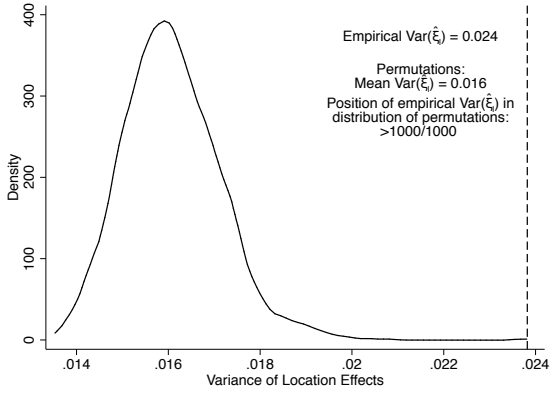
Permutation Test: 1,000 Random Reallocations of Plants

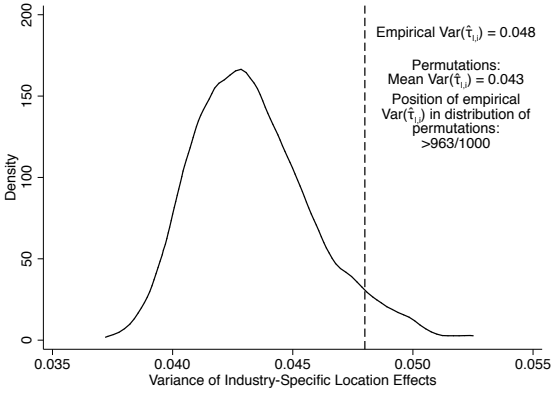


Permutation Test: 1,000 Random Reallocations of Plants

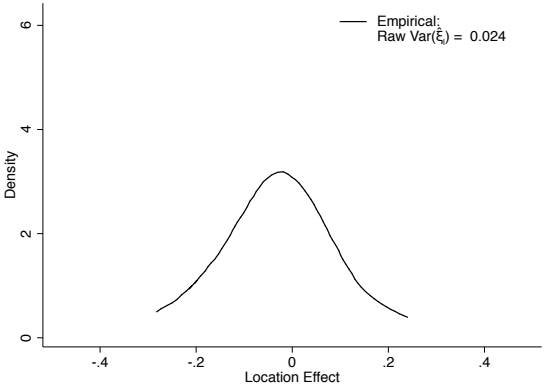


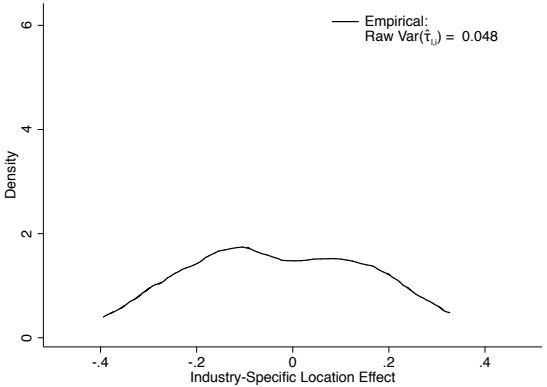
Permutation Test: 1,000 Random Reallocations of Plants



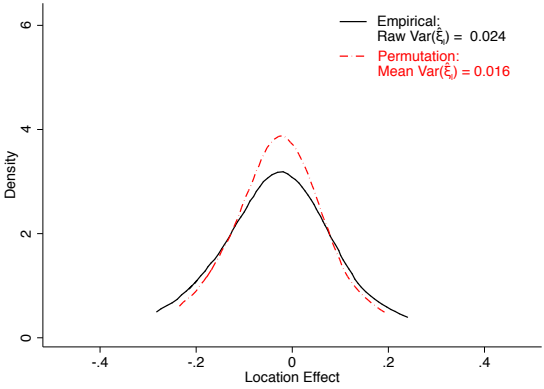


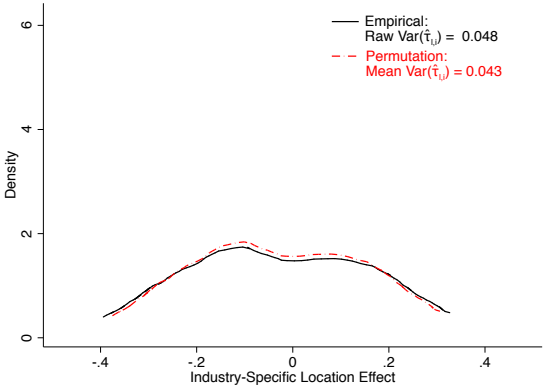
Taking Stock





Taking Stock





Outline

Definitions & statistical basics

Raw variance

Permutation test: granularity-bias-only benchmark

Bias correction: split-sample method

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 effects for l, $\widehat{\tau}_l^X, \widehat{\tau}_l^Y$

!!!! True place effect τ_l is common to both subsamples

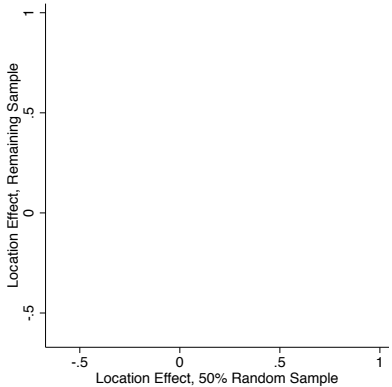
Covariance between subsamples is an unbiased estimator of variance of true place effects

$$\begin{aligned}\text{Cov}\left(\widehat{\tau}_l^X, \widehat{\tau}_l^Y, \right) &= \text{Cov}\left(\tau_l + u_l^{X_l}, \tau_l + u_l^{Y_l}\right) \\ &= \underbrace{\text{Var}\left(\tau_l\right)}_{\text{Var of True Place Effects}} + \underbrace{\text{Cov}\left(\tau_l, u_l^{X_l}\right)}_{=0} + \underbrace{\text{Cov}\left(\tau_l, u_l^{Y_l}\right)}_{=0} + \underbrace{\text{Cov}\left(u_l^{X_l}, u_l^{Y_l}\right)}_{=0}\end{aligned}$$

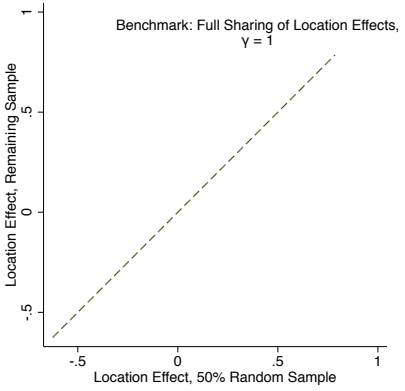
(where $u_l^{S_l} = \frac{1}{N^{S_l}} \sum_{p \in S_l} u_{pl}$)

In progress: shrinkage approach (e.g., Chetty, Friedman, Rockoff 2014)

Bias Correction of Variance: Split Samples

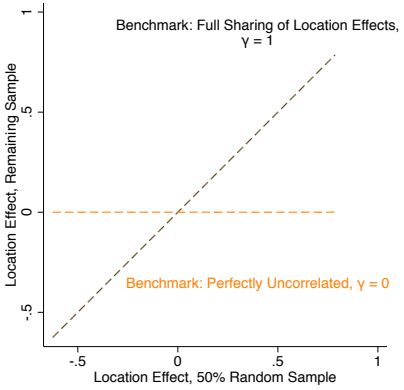


Bias Correction of Variance: Split Samples



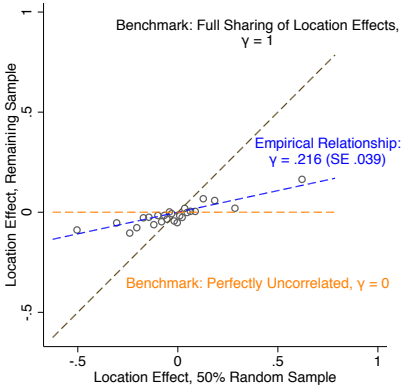
$$\gamma = \frac{\text{Cov}(\widehat{\tau}^Y, \widehat{\tau}^X)}{\text{Var}(\widehat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\widehat{\tau}^X)} = \text{i.e. the share of variance of true place effects buried in raw variance}$$

Bias Correction of Variance: Split Samples



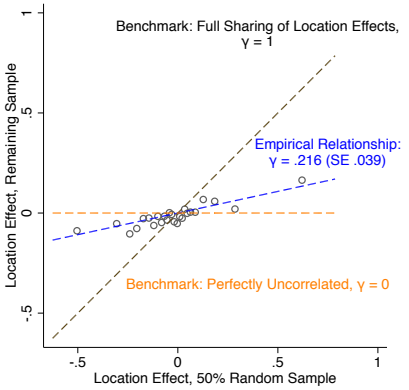
$$\gamma = \frac{\text{Cov}(\widehat{\tau}^Y, \widehat{\tau}^X)}{\text{Var}(\widehat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\widehat{\tau}^X)} = \text{i.e. the share of variance of true place effects buried in raw variance}$$

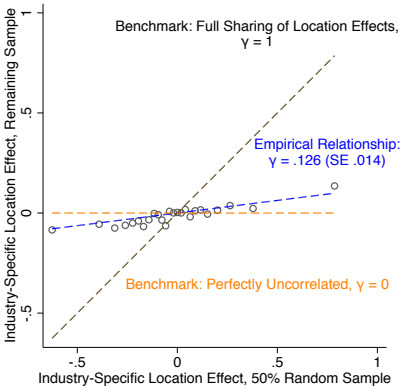
Bias Correction of Variance: Split Samples



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

Bias Correction of Variance: Split Samples

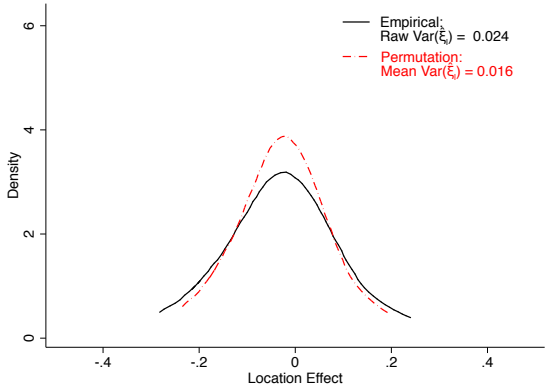


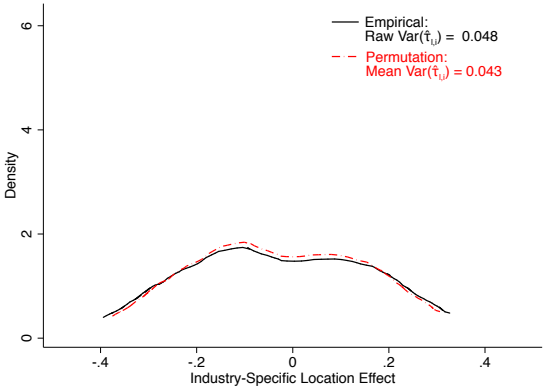


$$\gamma = \frac{\text{Cov}(\widehat{\tau}^Y, \widehat{\tau}^X)}{\text{Var}(\widehat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\widehat{\tau}^X)} = \text{i.e. the share of variance of true place effects buried in raw variance}$$

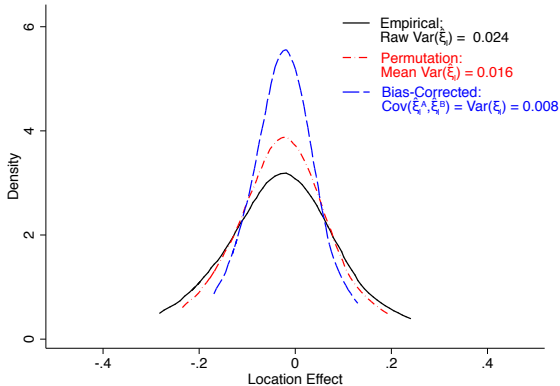
See paper: Sample split leads to higher raw variance by doubling granularity bias on x-axis, so γ is lower than the variance ratios on next slide using the full sample $\text{Var}(\widehat{\tau}^{X\,|\,Y})$ to compute raw variance .

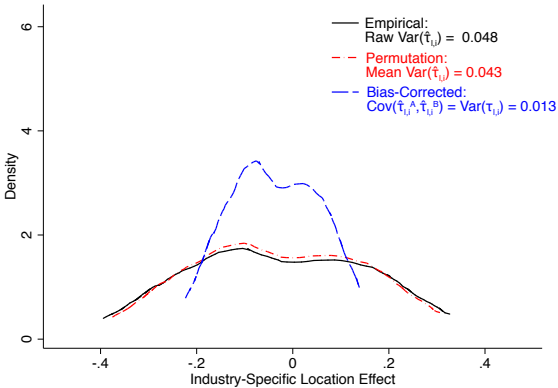
Bias Correction of Variance: Split Samples





Bias Correction of Variance: Split Samples





2/3 of raw variance is spurious – "granularity bias"

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

Outline

Definitions & statistical basics

Raw variance

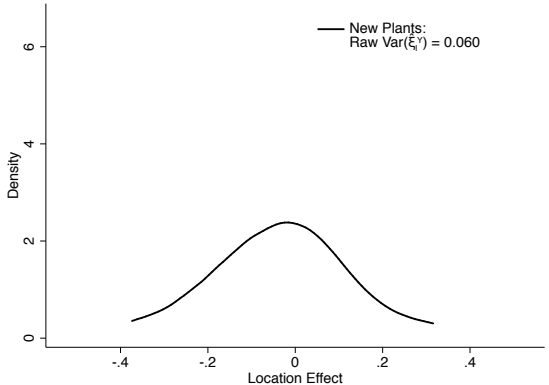
Permutation test: granularity-bias-only benchmark

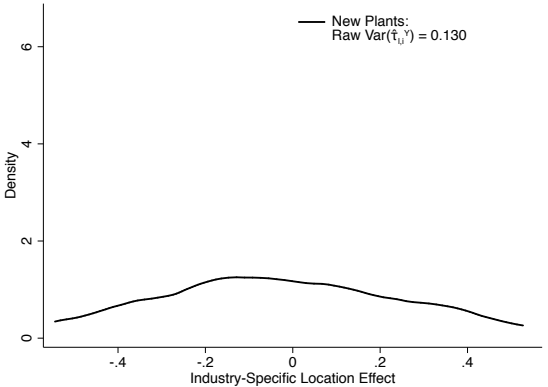
Bias correction: split-sample method

Extension I: new plants' place effects – 5 years and younger

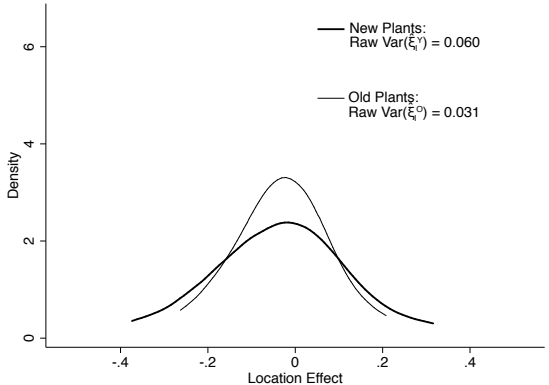
Extension II: within-country dispersion in 15 European countries

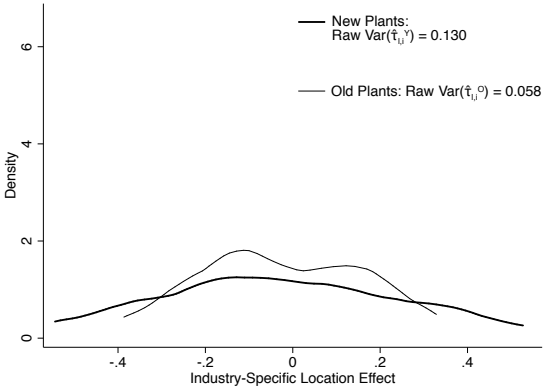
New Plants: Even Higher Raw Variance



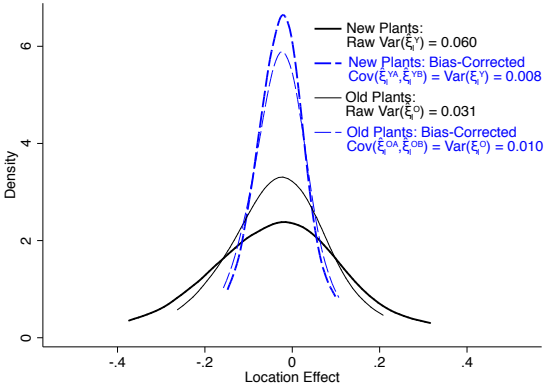


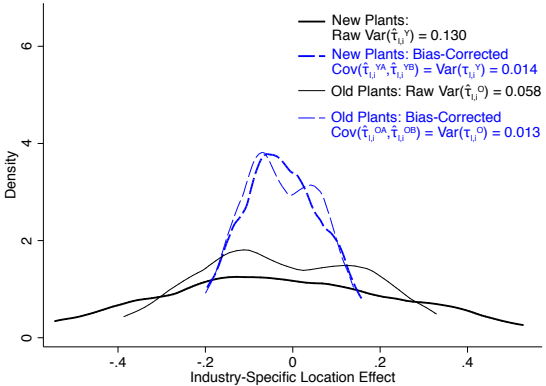
New Plants: Even Higher Raw Variance





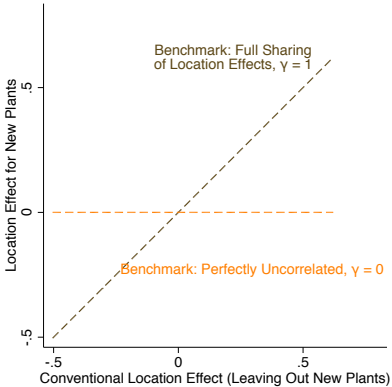
New Plants: Even Higher Bias

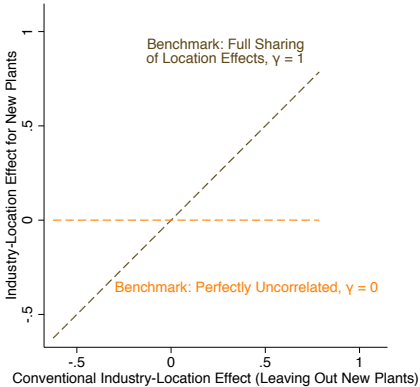




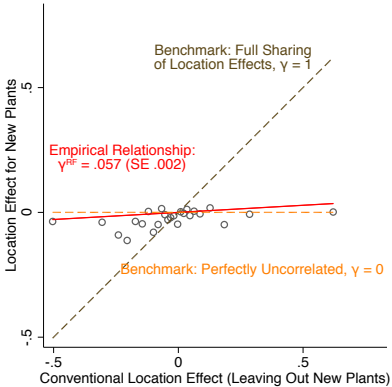
Bias corrected variances are very similar between new (0.008, 0.014) and old (0.010, 0.013)!

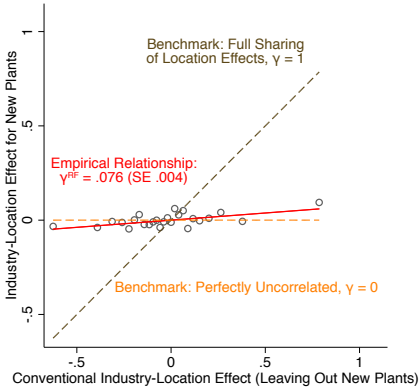
New Plants: Covariance With Old Plants



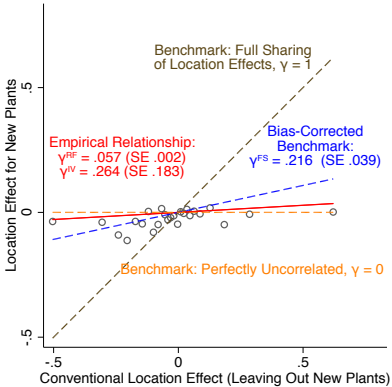


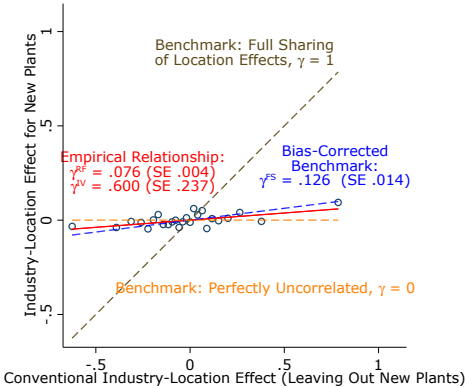
New Plants: Covariance With Old Plants





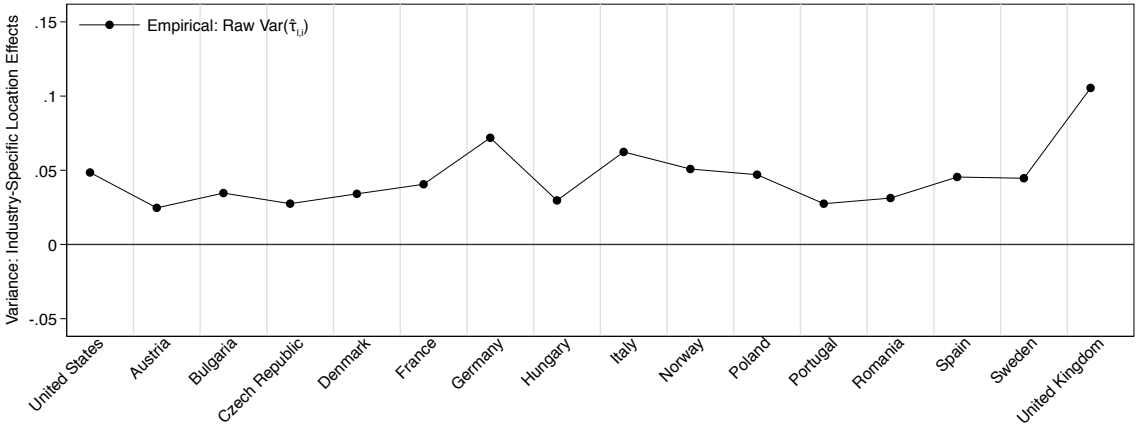
New Plants: Covariance With Old Plants





Outline

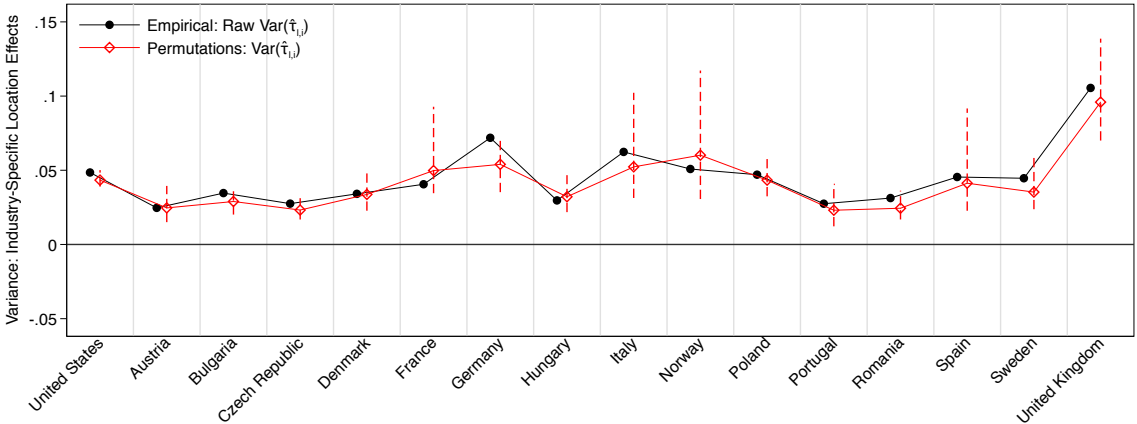
15 European Countries: Industry-Speci c Location E ects



European countries: NUTS-2 regions, 2-digit NACE industry, manufacturing sector. Bureau van Dijk rm data

USA: MSA, 4-digit NAICS, plant-level data (CMF)

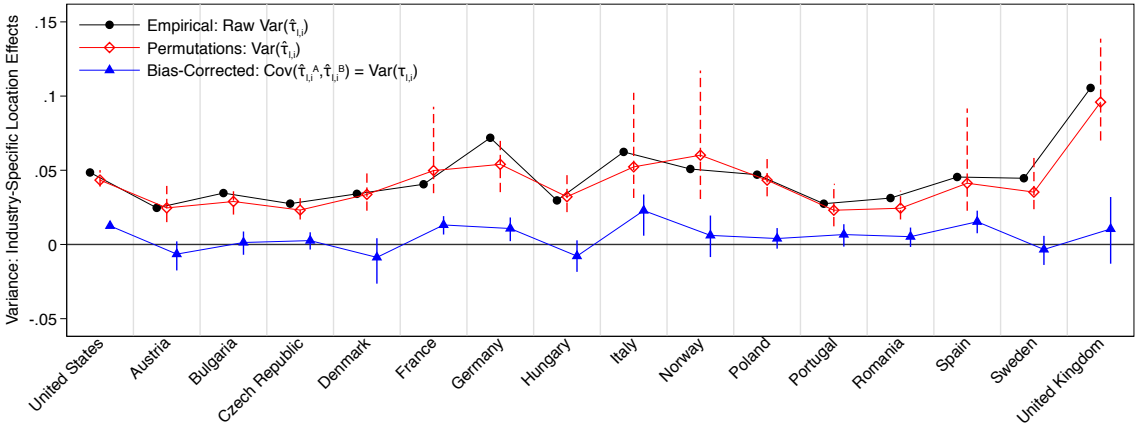
15 European Countries: Industry-Speci c Location E ects



European countries: NUTS-2 regions, 2-digit NACE industry, manufacturing sector. Bureau van Dijk rm data

USA: MSA, 4-digit NAICS, plant-level data (CMF)

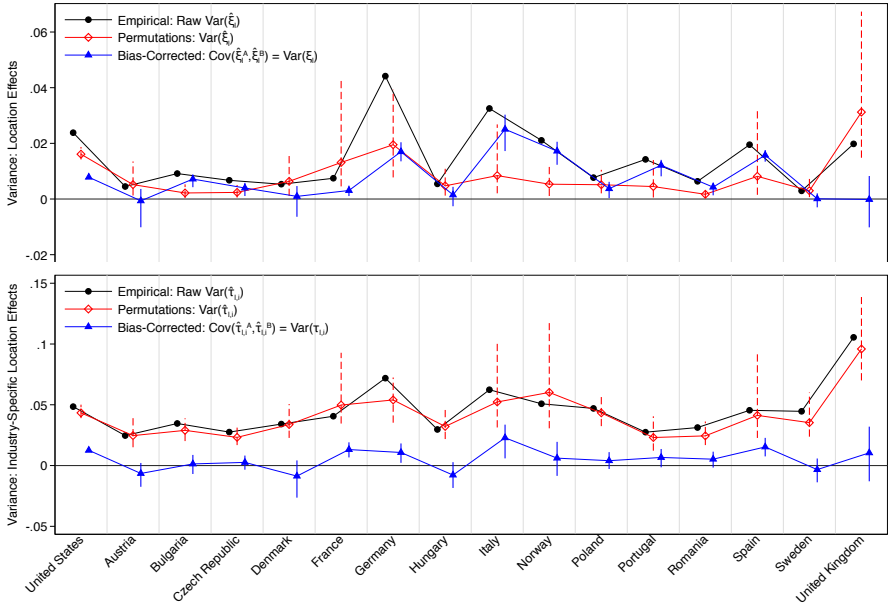
15 European Countries: Industry-Speci c Location E ects



European countries: NUTS-2 regions, 2-digit NACE industry, manufacturing sector. Bureau van Dijk rm data

USA: MSA, 4-digit NAICS, plant-level data (CMF)

15 European Countries: Location E ects



Measuring Var(Location-Speci c E[Plant-Level Productivity])

"True Place E ects"

Places do differ in plant productivity.

Large raw variance, but:

- 2/3 is spurious (granularity bias: idiosyncratic plant-level dispersion in productivity).
- 1/3 due to true place e ects.

Patterns extend to 15 European countries.

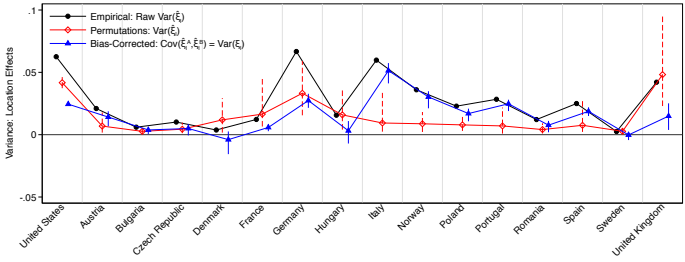
Bias larger for new plants.

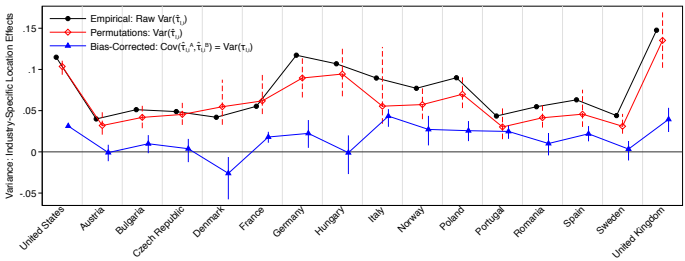
Place effects for new plants somewhat distinct from those of old plants.

Appendix Slides

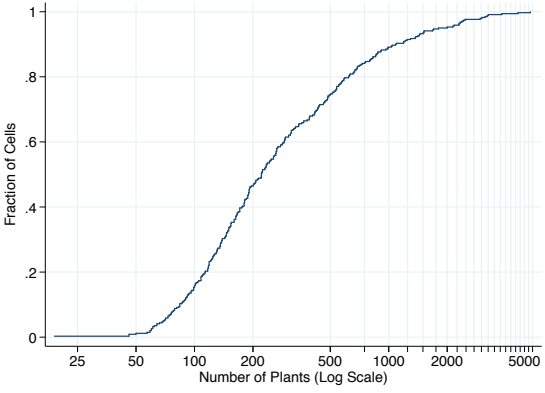
	All Plants					New Plants		Old Plants	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	TFPr	ln $\frac{\text{Vadd}}{\text{Emp}}$	unw'd	6-d	10 p	TFPr	ln $\frac{\text{VAdd}}{\text{Emp}}$	TFPr	ln $\frac{\text{VAdd}}{\text{Emp}}$
Panel A: Variance of Place Effects									
Raw var	0.024	0.063	0.012	0.019	0.017	0.060	0.139	0.031	0.081
Mean: perm's	0.016	0.042	(pending RDC reopening)			0.062	0.162	0.020	0.051
SD: perm's	0.001	0.002	(pending RDC reopening)			0.005	0.009	0.001	0.003
p-v	0.000	0.000	(pending RDC reopening)			0.006	0.088	0.000	0.000
SS 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 var	0.048	0.115	0.023	0.034	0.029	0.130	0.284	0.058	0.139
Mean: perm's	0.043	0.104	(pending RDC reopening)			0.128	0.303	0.052	0.121
SD: perm's	0.003	0.005	(pending RDC reopening)			0.008	0.013	0.003	0.006
p-v	0.037	0.034	(pending RDC reopening)			0.375	0.940	0.046	0.009
SS cov.	0.013	0.031	0.007	0.003	0.008	0.014	0.030	0.013	0.034
MSAs	(pending RDC reopening)								
IndXMSAs	(pending RDC reopening)								
Plants	(pending RDC reopening)								

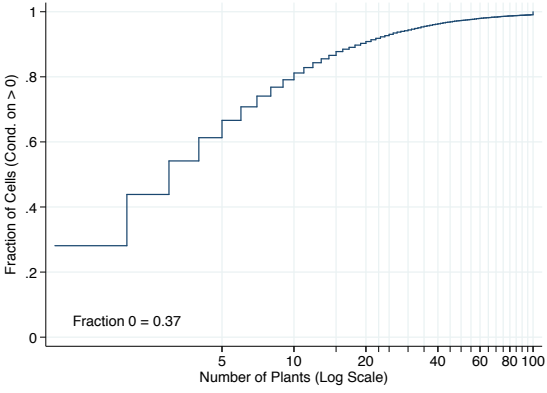
Log Value Added Per Worker





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)