Measuring Spatial Differences 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 finding:**

- 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

• 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:
  \[
  \tau_{l} = E[a_{pl} | l] = \int a dF_{s}^{l}(a) \Rightarrow a_{pl} = \tau_{l} + u_{pl}
  \]

- Measured average productivity of $N^{S_{l}}$ plants $p \in S_{l}$:
  \[
  \hat{\tau}_{l}^{S_{l}} = \frac{1}{N^{S_{l}}} \sum_{p \in S_{l}} a_{pl}.
  \]

- Of course, average $\hat{\tau}_{l}^{S_{l}}$ is an unbiased and consistent estimator of $\tau_{l}$…
Raw Variance of Place Averages

\[
\text{Var of Est. Place Effects (Location Averages)}
\]

\[
\text{Var}\left(\hat{\tau}_l^{S_l}\right) = \text{Var}\left(\frac{1}{N_{S_l}} \sum_{p \in S_l} a_{pl}\right)
\]

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

\[
= \text{Var}\left(\tau_l + \frac{1}{N_{S_l}} \sum_{p \in S_l} u_{pl}\right)
\]
Pitfalls: Granularity Bias

\[ \text{Var of Est. Place Effects (Location Averages)} \]

\[
\text{Var}(\widehat{\tau}_l^{S_i}) = \text{Var}\left( \frac{1}{N_{S_i}} \sum_{p \in S_i} a_{pl} \right)
\]

\[
= \text{Var}\left( \frac{1}{N_{S_i}} \sum_{p \in S_i} [\tau_l + u_{pl}] \right)
\]

\[
= \text{Var}\left( \tau_l + \frac{1}{N_{S_i}} \sum_{p \in S_i} u_{pl} \right)
\]

\[
= \text{Var}(\tau_l) + \frac{1}{N_{S_i}} \sum_{l \in L} \sigma_l(u)^2 + 2 \text{Cov} \left( \tau_l, \frac{1}{N_{S_i}} \sum_{p \in S_i} u_{pl} \right)
\]

>0 if \( N < \infty \land \sigma(u) > 0 \)

Bias from Granularity:
Var of Sample Means

Orthogonal by Construction

\( = 0 \)
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:
  \[ \hat{\tau}_{l(p),i(p)} = \text{Avg}[a_{pl|i,l}] - \bar{a}_i \]
  - Demeaned \(\Rightarrow\) "Average percent premium in TFP compared to national industry average"
  - 4-digit NAICS x MSA [Robustness: 6-digit]
  - \(\geq 5\) plants per cell [Robustness: at least \(\geq 10\)]

- Overall location effect:
  \[ \hat{\xi}_{l(p)} = \text{Avg}[\hat{\tau}_{l(p),i(p)}|l] \]
  - Location average of its industry premia \(\hat{\tau}_{l(p),i(p)}\) [Robustness: unweighted]
  - Weighting: plant employment [Robustness: log value added per worker]
  - Main measure: TFPr (follow Foster et al. 2008)

- Estimate via fixed effects regressions

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

Empirical:

Raw $\text{Var}(\hat{\xi}) = 0.024$
Raw Place Effects

Empirical:
Raw $\text{Var}(\xi) = 0.024$

Empirical:
Raw $\text{Var}(\tau_{l,i}) = 0.048$
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

\[
\text{Var}(\hat{\tau}_{Sl}) = \text{Var}(\tau) + \frac{1}{L} \sum_{l \in L} \frac{\sigma_l(u)^2}{NS_l} + 2 \text{Cov} \left( \tau, \frac{1}{NS_l} \sum_{p \in S_l} a_{pl} \right)
\]

= 0 if \(N < \infty \land \sigma(a) > 0\)

Bias from Granularity: Var of Sample Means

Orthogonal by Construction

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

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

\[
\Rightarrow \text{Var}(\hat{\tau}_{Sl}) = \text{Var}(\tau) + \sigma(a)^2 \cdot \sum_{l \in L} \frac{1}{NS_l L} + 2 \text{Cov} \left( \tau, \frac{1}{NS_l} \sum_{p \in S_l} a_{pl} \right)
\]

= 0 if \(N < \infty \land \sigma(a) > 0\)

Bias from Granularity: Var of Sample Means

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

![Graph showing variance of location effects against density]
Permutation Test: 1,000 Random Reallocations of Plants
Permutation Test: 1,000 Random Reallocations of Plants

Empirical $\text{Var}(\hat{\xi}_l) = 0.024$

Permutations:
Mean $\text{Var}(\hat{\xi}_l) = 0.016$
Position of empirical $\text{Var}(\hat{\xi}_l)$ in distribution of permutations:
$>1000/1000$
Permutation Test: 1,000 Random Reallocations of Plants

Empirical \text{Var}(\hat{\xi}) = 0.024

Permutations:
Mean \text{Var}(\hat{\xi}) = 0.016
Position of empirical \text{Var}(\hat{\xi}) in distribution of permutations:
>1000/1000

Empirical \text{Var}(\hat{\tau}_{li}) = 0.048

Permutations:
Mean \text{Var}(\hat{\tau}_{li}) = 0.043
Position of empirical \text{Var}(\hat{\tau}_{li}) in distribution of permutations:
>963/1000
Taking Stock

Empirical:
Raw Var(\(\xi_l\)) = 0.024

Empirical:
Raw Var(\(\tau_{l,i}\)) = 0.048
Taking Stock

Empirical:
\[ \text{Raw Var}(\xi) = 0.024 \]

Permutation:
\[ \text{Mean Var}(\xi) = 0.016 \]

Empirical:
\[ \text{Raw Var}(\tau_{l,i}) = 0.048 \]

Permutation:
\[ \text{Mean Var}(\tau_{l,i}) = 0.043 \]
• 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 $l$, $\hat{\tau}_l^X, \hat{\tau}_l^Y$

True place effect $\tau_l$ is common to both subsamples

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

\[
\text{Cov}(\hat{\tau}_l^X, \hat{\tau}_l^Y) = \text{Cov}(\tau_l + \bar{u}_l^X, \tau_l + \bar{u}_l^Y) \\
= \text{Var}(\tau_l) + \text{Cov}(\tau_l, \bar{u}_l^X) + \text{Cov}(\tau_l, \bar{u}_l^Y) + \text{Cov}(\bar{u}_l^X, \bar{u}_l^Y) \\
= \text{Var of True Place Effects} + 0 + 0 + 0
\]

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

In progress: shrinkage approach (e.g., Chetty, Friedman, Rockoff 2014)
Bias Correction of Variance: Split Samples
Bias Correction of Variance: Split Samples

\[ \gamma = \frac{\text{Cov}(\hat{\tau}^Y, \hat{\tau}^X)}{\text{Var}(\hat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\hat{\tau}^X)} = \text{i.e. the share of variance of true place effects buried in raw variance} \]
Bias Correction of Variance: Split Samples

\[
\gamma = \frac{\text{Cov}(\hat{\tau}_Y, \hat{\tau}_X)}{\text{Var}(\hat{\tau}_Y)} = \frac{\text{Var}(\tau)}{\text{Var}(\hat{\tau}_X)} = \text{i.e. the share of variance of true place effects buried in raw variance}
\]
Bias Correction of Variance: Split Samples

\[ \gamma = \frac{\text{Cov}(\hat{\tau}^Y, \hat{\tau}^X)}{\text{Var}(\hat{\tau}^X)} = \frac{\text{Var}(\tau)}{\text{Var}(\hat{\tau}^X)} = \text{i.e. the share of variance of true place effects buried in raw variance} \]
Bias Correction of Variance: Split Samples

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

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

Empirical:
\[ \text{Raw Var}(\xi^l) = 0.024 \]

Permutation:
\[ \text{Mean Var}(\xi^l) = 0.016 \]

Empirical:
\[ \text{Raw Var}(\tau^l,i) = 0.048 \]

Permutation:
\[ \text{Mean Var}(\tau^l,i) = 0.043 \]
Bias Correction of Variance: Split Samples

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

\[ \Rightarrow \text{At most } 1/3 \text{ reflects variance of true place effects} \]
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:                              Raw Var(\(\xi_l Y\)) = 0.060

New Plants:                                   Raw Var(\(\tau_{l,i} Y\)) = 0.130
New Plants: Even Higher Raw Variance

New Plants:
Raw Var(\(\xi_Y\)) = 0.060

Old Plants:
Raw Var(\(\xi_O\)) = 0.031

Industry-Specific Location Effect

New Plants:
Raw Var(\(\tau_{l,i}^Y\)) = 0.130

Old Plants: Raw Var(\(\tau_{l,i}^O\)) = 0.058
New Plants: Even Higher Bias

New Plants: Bias-Corrected
\[ \text{Cov}(\hat{\xi}_{lY}, \hat{\xi}_{lY}) = \text{Var}(\xi_{lY}) = 0.008 \]

Old Plants: Bias-Corrected
\[ \text{Cov}(\hat{\xi}_{lO}, \hat{\xi}_{lO}) = \text{Var}(\xi_{lO}) = 0.010 \]

Bias corrected variances are very similar between new (0.008, 0.014) and old (0.010, 0.013)!
New Plants: Covariance With Old Plants

Benchmark: Perfectly Uncorrelated, $\gamma = 0$

Benchmark: Full Sharing of Location Effects, $\gamma = 1$

Location Effect for New Plants

Conventional Location Effect (Leaving Out New Plants)

Industry-Location Effect for New Plants

Conventional Industry-Location Effect (Leaving Out New Plants)
New Plants: Covariance With Old Plants

Empirical Relationship:
\[ \gamma_{RF} = 0.057 \pm 0.002 \]

Benchmark: Perfectly Uncorrelated, \( \gamma = 0 \)
Benchmark: Full Sharing of Location Effects, \( \gamma = 1 \)

Location Effect for New Plants

Conventional Location Effect (Leaving Out New Plants)

Empirical Relationship:
\[ \gamma_{RF} = 0.076 \pm 0.004 \]

Benchmark: Perfectly Uncorrelated, \( \gamma = 0 \)
Benchmark: Full Sharing of Location Effects, \( \gamma = 1 \)

Industry-Location Effect for New Plants

Conventional Industry-Location Effect (Leaving Out New Plants)
New Plants: Covariance With Old Plants

Benchmark: Perfectly Uncorrelated, $\gamma = 0$

Benchmark: Full Sharing of Location Effects, $\gamma = 1$

Bias-Corrected Benchmark: $\gamma_{FS} = 0.126$ (SE 0.014)

Empirical Relationship:
- $\gamma_{RF} = 0.076$ (SE 0.004)
- $\gamma_{IV} = 0.600$ (SE 0.237)

Location Effect for New Plants

- Conventional Location Effect (Leaving Out New Plants)

Industry-Location Effect for New Plants

- Conventional Industry-Location Effect (Leaving Out New Plants)
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
15 European Countries: Industry-Specific Location Effects

European countries: NUTS-2 regions, 2-digit NACE industry, manufacturing sector. Bureau van Dijk firm data
USA: MSA, 4-digit NAICS, plant-level data (CMF)
European countries: NUTS-2 regions, 2-digit NACE industry, manufacturing sector. Bureau van Dijk firm data
USA: MSA, 4-digit NAICS, plant-level data (CMF)
European countries: NUTS-2 regions, 2-digit NACE industry, manufacturing sector. Bureau van Dijk firm data
USA: MSA, 4-digit NAICS, plant-level data (CMF)
15 European Countries: Location Effects

Variance: Location Effects
- United States
- Austria
- Bulgaria
- Czech Republic
- Denmark
- France
- Germany
- Hungary
- Italy
- Norway
- Poland
- Portugal
- Romania
- Spain
- Sweden
- United Kingdom

Empirical: $\text{Raw } \text{Var}(\xi)$
Permutations: $\text{Var}(\xi)$
Bias-Corrected: $\text{Cov}(\hat{\xi}_A, \hat{\xi}_B) = \text{Var}(\xi)$

Variance: Industry-Specific Location Effects
- United States
- Austria
- Bulgaria
- Czech Republic
- Denmark
- France
- Germany
- Hungary
- Italy
- Norway
- Poland
- Portugal
- Romania
- Spain
- Sweden
- United Kingdom

Empirical: $\text{Raw } \text{Var}(\tau_{l,i})$
Permutations: $\text{Var}(\tau_{l,i})$
Bias-Corrected: $\text{Cov}(\hat{\tau}_{l,iA}, \hat{\tau}_{l,iB}) = \text{Var}(\tau_{l,i})$
Measuring $\text{Var}(\text{Location-Specific } E[\text{Plant-Level Productivity}])$

"True Place Effects"

- Places do differ in plant productivity.

- Large raw variance, but:
  - 2/3 is spurious (granularity bias: idiosyncratic plant-level dispersion in productivity).
  - $\Leftrightarrow$ 1/3 due to true place effects.

- 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
<table>
<thead>
<tr>
<th></th>
<th>All Plants</th>
<th>New Plants</th>
<th>Old Plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) TFPr</td>
<td>(2) ln Vadd_Emp</td>
<td>(3) unw’d 6-d</td>
<td>(4) ≤ 10 p</td>
</tr>
<tr>
<td><strong>Panel A: Variance of Place Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw var</td>
<td>0.024</td>
<td>0.063</td>
<td>0.012</td>
</tr>
<tr>
<td>Mean: perm’s</td>
<td>0.016</td>
<td>0.042</td>
<td>(pending RDC reopening)</td>
</tr>
<tr>
<td>SD: perm’s</td>
<td>0.001</td>
<td>0.002</td>
<td>(pending RDC reopening)</td>
</tr>
<tr>
<td>p-v</td>
<td>0.000</td>
<td>0.000</td>
<td>(pending RDC reopening)</td>
</tr>
<tr>
<td>SS cov.</td>
<td>0.008</td>
<td>0.025</td>
<td>0.005</td>
</tr>
<tr>
<td><strong>Panel B: Variance of Place-Industry Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw var</td>
<td>0.048</td>
<td>0.115</td>
<td>0.023</td>
</tr>
<tr>
<td>Mean: perm’s</td>
<td>0.043</td>
<td>0.104</td>
<td>(pending RDC reopening)</td>
</tr>
<tr>
<td>SD: perm’s</td>
<td>0.003</td>
<td>0.005</td>
<td>(pending RDC reopening)</td>
</tr>
<tr>
<td>p-v</td>
<td>0.037</td>
<td>0.034</td>
<td>(pending RDC reopening)</td>
</tr>
<tr>
<td>SS cov.</td>
<td>0.013</td>
<td>0.031</td>
<td>0.007</td>
</tr>
</tbody>
</table>

MSAs (pending RDC reopening)  
IndXMSAs (pending RDC reopening)  
Plants (pending RDC reopening)
Log Value Added Per Worker

Variance: Location Effects

- United States
- Austria
- Bulgaria
- Czech Republic
- Denmark
- France
- Germany
- Hungary
- Italy
- Norway
- Poland
- Portugal
- Romania
- Spain
- Sweden
- United Kingdom

Empirical: Raw Var(ξ̂)
Permutations: Var(ξ̂)
Bias-Corrected: Cov(ξ̂A,ξ̂B) = Var(ξ̂)

Variance: Industry-Specific Location Effects

- United States
- Austria
- Bulgaria
- Czech Republic
- Denmark
- France
- Germany
- Hungary
- Italy
- Norway
- Poland
- Portugal
- Romania
- Spain
- Sweden
- United Kingdom

Empirical: Raw Var(τ̂l,i)
Permutations: Var(τ̂l,i)
Bias-Corrected: Cov(τ̂l,iA,τ̂l,iB) = Var(τl,i)
Cell Counts

Fraction of Cells (Cond. on > 0) vs. Number of Plants (Log Scale)

Fraction 0 = 0.37