Spatial economics for granular settings

Jonathan I. Dingel (Chicago Booth)
Felix Tintelnot (UChicago)

December 2020

Quantitative spatial models in granular settings

- Spatial linkages (commuting, trade, local externalities, etc) govern the incidence of local economic shocks
- Want "an empirically relevant quantitative model to perform general equilibrium counterfactual policy exercises" (Redding and Rossi-Hansberg, 2017)
- Continuum of agents → observed shares = model probabilities ▸ Literature
- High-resolution spatial settings are **granular**: an individual decision maker is large relative to the economic outcome examined
- Challenges for producing predictions in granular settings:
  - Estimation: is an outcome twice as probable because two people chose it?
  - Theory: individual choices affect local labor supply and land demand
  - Counterfactuals: equilibrium outcomes depend on individual idiosyncrasies

Spatial economics for granular settings: Roadmap

Computing counterfactuals in continuum models

Counterfactual analysis in granular empirical settings

- Apply continuum model to NYC 2010
- Monte Carlo: Calibrated-shares procedure overfits data
- Event studies: Neighborhood employment booms

Granular model

Application to Amazon's HQ2

Computing counterfactual outcomes in continuum models

Continuum model: Economic environment

- Each location has productivity A and land endowment T
- Measure L individuals w/ one unit of labor and hired by competitive firms producing freely traded goods differentiated by location of production
- Individuals have Cobb-Douglas preferences over goods $(1-\alpha)$ and land (α)
- Individuals have idiosyncratic tastes for pairs of residential and workplace locations, such that i's utility from living in k and working in n is

$$U_{kn}^{i} = \epsilon \ln \left(\frac{w_n}{r_k^{\alpha} P^{1-\alpha} \delta_{kn}} \right) + \nu_{kn}^{i} \qquad \nu_{kn}^{i} \overset{\text{iid}}{\sim} \text{T1EV}$$

Continuum model: Equilibrium

Given economic primitives (α, ϵ, σ,L,$\{A_n\}$,$\{T_k\}$,$\{\delta_{kn}\}$), an equilibrium is a set of wages $\{w_n\}$, rents $\{r_k\}$, and labor allocation $\{\ell_{kn}\}$ such that

$$\text{labor allocation:} \qquad \frac{\ell_{kn}}{L} = \frac{w_n^{\epsilon}\left(r_k^{\alpha}\delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon}\left(r_{k'}^{\alpha}\delta_{k'n'}\right)^{-\epsilon}} \tag{1}$$

$$\text{goods markets:} \qquad A_n \sum_k \frac{\ell_{kn}}{\delta_{kn}} = \frac{\left(w_n/A_n\right)^{-\sigma}}{P^{1-\sigma}} Y \qquad \forall n \tag{2}$$

$$\text{land markets:} \qquad T_k = \frac{\alpha}{r_k} \sum_n \frac{\ell_{kn}}{\delta_{kn}} w_n \qquad \forall k \tag{3}$$

$\left(\frac{1+\epsilon}{\sigma+\epsilon}\right)\left(\frac{\alpha\epsilon}{1+\alpha\epsilon}\right) \leq \frac{1}{2} \implies$ unique equilibrium (Allen, Arkolakis and Li, 2020)

Continuum model: Procedures for counterfactual predictions

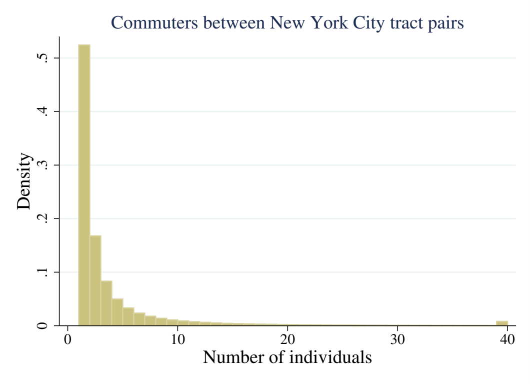
1. Covariates-based approach (e.g., Ahlfeldt et al. 2015)
    - Parameterize δ_{kn} as function of observed covariates
    - After estimating model, compute outcomes at counterfactual values
    - Equation (1) generically not satisfied by observed ℓ_{kn} at chosen δ_{kn}
2. Calibrated-shares procedure ("exact hat algebra" from trade) ▸ Math
    - Infer combinations of $(\{A_n\},\{T_k\},\{\delta_{kn}\})$ by assuming equation (1) satisfied by observed ℓ_{kn} and w_n (e.g., $\ell_{kn} = 0 \implies \delta_{kn} = \infty$)
    - Compute counterfactual outcomes due to proportionate changes in $\{A_n\}$, $\{T_k\}$, or $\{\delta_{kn}\}$ (without knowing initial levels)
    - Used far more frequently than the covariates-based approach

Counterfactual analysis in granular empirical settings

NYC is a granular setting

NYC has 2.5 million resident-employees and 4.6 million tract pairs.

- 84% of tract pairs have zero commuters between them
- 40.7% of commuters in cell with ≤ 5
- 44% of NYC tract pairs with positive flow in 2013 were zeros in 2014
- Gravity model predicts 2014 value better than 2013 value for bottom 95% of tract pairs ▸



Source: Longitudinal Employer-Household Dynamics, Origin Destination Employment Statistics. LODES employment counts are noise-infused and LODES flows are synthetically generated.

▸ Detroit ▸ MSP ▸ Granular counties ▸ Impersistent: Detroit ▸ Imp: MSP ▸ Imp: counties ▸ Asymmetric zeros

Parameterization of commuting costs

- Pick $\alpha = 0.24$, $\sigma = 4$, $L =$ number of employed individuals
- Seek values of $\{\delta_{kn}\}$, ϵ, $\{T_k\}$, $\{A_n\}$

$$\delta_{kn} = \underbrace{\bar{\delta}_{kn}}_{\text{observed}} \times \underbrace{\lambda_{kn}}_{\text{unobserved}}$$

- Compute $\{\bar{\delta}_{kn}\}$ from Google Maps transit times: $\bar{\delta}_{kn} = \frac{H}{H - t_{kn} - t_{nk}}$
    1. Covariates-based approach: Assume $\lambda_{kn} = 1 \; \forall k, n$
    2. Calibrated-shares procedure: Assume structural error λ_{kn} appropriately orthogonal

Estimating the commuting elasticity for NYC in 2010

Logit log likelihood function

$$\ln \mathcal{L} = \sum_k \sum_n \ell_{kn} \ln \left[\frac{w_n^{\epsilon} \left(r_k^{\alpha} \bar{\delta}_{kn} \right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon} \left(r_{k'}^{\alpha} \bar{\delta}_{k'n'} \right)^{-\epsilon}} \right]$$

Commuting gravity equation

$$\frac{\ell_{kn}}{L} = \frac{w_n^{\epsilon} \left(r_k^{\alpha} \bar{\delta}_{kn} \lambda_{kn} \right)^{-\epsilon}}{\sum_{k',n'} w_{n'}^{\epsilon} \left(r_{k'}^{\alpha} \bar{\delta}_{k'n'} \lambda_{k'n'} \right)^{-\epsilon}}$$

	NYC (2010) MLE
Commuting cost	-7.986
	(0.307)
Model fit (pseudo-R^2)	0.662
Location pairs	4,628,878
Commuters	2,488,905

NOTES: Specification includes residence fixed effects and workplace fixed effects.

Covariates-based approach: Solve for $\{T_k\}$ and $\{A_n\}$ using fixed effects ($\propto r_k^{-\alpha\epsilon}$ and w_n^{ϵ}) and equations (1), (2), and (3)

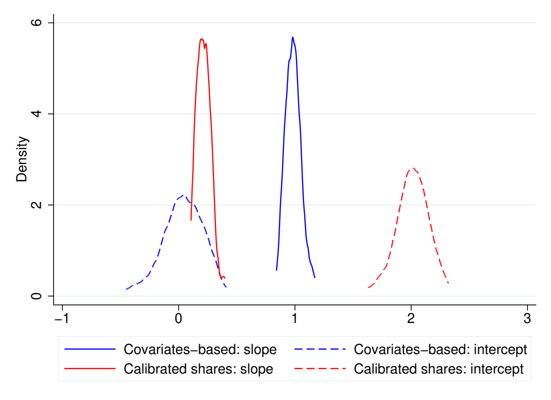
Calibrated-shares procedure: Use estimated ϵ

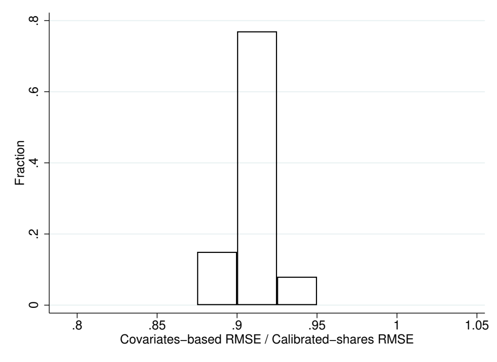
Monte Carlo: Applying each procedure to granular data

- DGP is estimated covariates-based model for NYC in 2010
- Simulated "event": $\uparrow$ productivity of 200 Fifth Ave tract by 18%
- 100 simulations of 2.5 million draws from ex ante and ex post data-generating process (interpreting ℓ_{kn}/L as probability)
- Apply calibrated-shares procedure and covariates-based approach (Increase A_n to match total employment increase in simulated data)
- Does the procedure predict the change in the number of commuters from each residential tract working in the "treated" tract?
- Regress "observed" changes on predicted changes (2160 obs per simulation)
- Ideally, want slope $= 1$ and intercept $= 0$
- Compute forecast errors (RMSE for "observed" vs predicted changes)

Monte Carlo: Calibrated-shares procedure performs poorly

Apply each procedure to simulated "2010" & "2012" data. 100 simulations w/ $I = 2,488,905$





I	2.5	5	12.5	25	50	125	250	2560
slope	0.22	0.55	0.82	0.90	0.95	0.98	0.99	1.00
intercept	2.01	1.15	0.47	0.25	0.12	0.05	0.03	0.00

Using tract-level events to evaluate model performance

Kehoe (2005): "it is the responsibility of modelers to demonstrate that their models are capable of predicting observed changes, at least ex post"

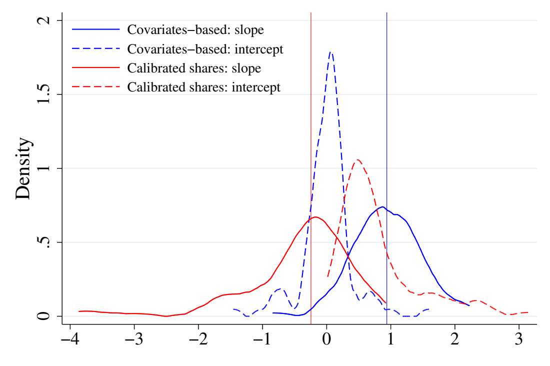
How well do models predict changes in commuting flows?

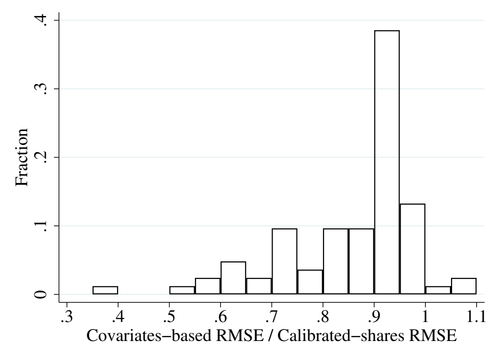
- Look at 83 tract-level employment booms (+12.5%) in NYC in 2010–2012

  e.g., Tiffany & Co. moving to 200 Fifth Avenue and Google moving to 111 Eighth Avenue
- We raise productivity in tract to match observed change in total employment
- Does the model predict changes in bilateral commuting flows to that destination? (n.b. total employment change need not be exogenous)
  - Regress observed changes on predicted changes
  - Ideally, want slope = 1 and intercept = 0

Comparison of models' predictive performance across 83 events

Covariates-based model much better at predicting change in number of commuters from each residential tract to booming workplace tract





▸ Pooled data ▸ No extensive margin

A quantitative spatial model for granular settings

A granular quantitative spatial model

- We introduce a granular model with an integer number of individuals
- In the limit ($I \to \infty$), our model is the standard quantitative spatial model
- For now, skip bells and whistles to focus on granular vs continuum cases

Modeling granularity:

- Individuals must have beliefs about equilibrium wages and land prices

$$\binom{I + N^2 - 1}{N^2 - 1} = \frac{(I + N^2 - 1)!}{(N^2 - 1)!I!} \qquad I = 10, N = 4 \implies 3.27 \times 10^6$$

- There will be a *distribution* of equilibria for each set of parameters

Model: Economic environment

- Each location has productivity A and land endowment T
- I individuals are endowed with L/I units of labor and hired by competitive firms producing freely traded goods differentiated by location of production
- Individuals have Cobb-Douglas preferences over goods and land
- Individuals have idiosyncratic tastes for residence-workplace pairs
- Workers know primitives (α, ϵ, σ,I,L,$\{A_n\}$,$\{T_k\}$,$\{\delta_{kn}\}$) and have (common) point-mass beliefs $\tilde{r}_k$ and $\tilde{w}_n$ about land prices and wages
- Worker i knows idiosyncratic preferences $\{\nu^i_{kn}\}$

Timing: Individuals choose labor allocation, then markets clear

1. Workers choose the kn pair that maximizes

$$\tilde{U}^i_{kn} = \epsilon \ln \left(\frac{\tilde{w}_n}{\tilde{P}^{1-\alpha} \tilde{r}_k^{\alpha} \delta_{kn}} \right) + \nu^i_{kn}$$

   given point-mass beliefs $\tilde{r}_k$ and $\tilde{w}_n$

2. After choosing kn based on their beliefs, workers are immobile and cannot relocate
3. Given the labor allocation $\{\ell_{kn}\}$, **a trade equilibrium** is a set of wages $\{w_n\}$ and land prices $\{r_k\}$ that clears all markets.

Concept: Granular commuting equilibrium

Given belief vectors $\{\tilde{w}_n\}$ and $\{\tilde{r}_k\}$, logit probabilities for kn pairs:

$$\Pr(U^i_{kn} > U^i_{k'n'}\ \forall (k', n') \neq (k, n)) = \frac{\tilde{w}_n^{\epsilon} \left(\tilde{r}_k^{\alpha} \delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}_{n'}^{\epsilon} \left(\tilde{r}_{k'}^{\alpha} \delta_{k'n'}\right)^{-\epsilon}}. \tag{4}$$

Given primitives (α, ϵ, σ,I,L,$\{A_n\}$,$\{T_k\}$,$\{\delta_{kn}\}$) and point-mass beliefs $\{\tilde{w}_n\}, \{\tilde{r}_k\}$, a **granular commuting equilibrium** is defined as a labor allocation $\{\ell_{kn}\}$, wages $\{w_n\}$, and land prices $\{r_k\}$ such that

- $\{\ell_{kn}\}$ is the labor allocation resulting from I independent draws from the probability function in equation (4); and
- wages $\{w_n\}$ and land prices $\{r_k\}$ are *a trade equilibrium* given the labor allocation $\{\ell_{kn}\}$.

Limit of granular commuting equilibrium is continuum equilibrium

- Aggregate labor supply L is fixed. Taking the limit $I \to \infty$ yields

$$\frac{\ell_{kn}}{L} = \frac{\tilde{w}_n^{\epsilon} \left(\tilde{r}_k^{\alpha} \delta_{kn} \right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}_{n'}^{\epsilon} \left(\tilde{r}_{k'}^{\alpha} \delta_{k'n'} \right)^{-\epsilon}}. \tag{5}$$

- Definition: $\tilde{w}$ and $\tilde{r}$ are "**continuum-case rational expectations**" if $\tilde{w}$ and $\tilde{r}$ constitute a trade equilibrium for the labor allocation $\{\ell_{kn}\}$ given by equation (5).
- Result: As $I \to \infty$, if individuals' point-mass beliefs are continuum-case rational expectations, then the granular model's equilibrium quantities and prices coincide with those of the continuum model.

Estimating the granular model

Granular model's likelihood (McFadden, 1974, 1978; Guimarães, Figueirdo and Woodward, 2003)

$$\ln \mathcal{L} = \sum_k \sum_n \ell_{kn} \ln \left[\frac{\tilde{w}_n^{\epsilon} \left(\tilde{r}_k^{\alpha} \bar{\delta}_{kn} \right)^{-\epsilon}}{\sum_{k',n'} \tilde{w}_{n'}^{\epsilon} \left(\tilde{r}_{k'}^{\alpha} \bar{\delta}_{k'n'} \right)^{-\epsilon}} \right]$$

- Solve for $\{T_k\}$ and $\{A_n\}$ using fixed effects ($\propto \tilde{r}_k^{-\alpha\epsilon}$ and $\tilde{w}_n^{\epsilon}$) under CCRE
- This estimation procedure yields same ϵ, $\{T_k\}$, and $\{A_n\}$ as the covariates-based continuum model

Ex post regret is small

- Individuals make residence-workplace choices based on wage and rent beliefs
- The realized equilibrium wages and rents will differ ‣ Price dispersion
- Calculate ex post regret for kn at realized prices:

$$\frac{\max_{k',n'} U^i_{k',n'}}{U^i_{k,n}} - 1 = \frac{\max_{k',n'} \left(\epsilon \ln \left(\frac{w_{n'}}{P^{1-\alpha} r^{\alpha}_{k'} \delta_{k'n'}}\right) + \nu^i_{k'n'}\right)}{\left(\epsilon \ln \left(\frac{w_n}{P^{1-\alpha} r^{\alpha}_{k} \delta_{kn}}\right) + \nu^i_{kn}\right)} - 1$$

- Quantitatively modest: 96% would not want to switch ‣ Switchers
- Utility gain for median switcher would be 0.18% (1.36% for 99$^{\text{th}}$)

‣ Price dispersion in ex post regret simulations

Computing counterfactual outcomes using granular model

Continuum model's $\frac{\ell_{kn}}{L} = \mathbb{E}\left[\Pr(U^i_{kn} > U^i_{k'n'}\ \forall (k',n') \neq (k,n))\right]$, so (mean) quantities coincide

Granular uncertainty: individual idiosyncrasies $\rightarrow$ distributions of equilibrium quantities and prices

Compute confidence interval for change in residents in k: $\sum_n \ell'_{kn} - \sum_n \ell_{kn}$

- Characterize by simulations of granular model
- Normal approximation of binomial distribution for quantities

$$\text{std dev}\left(\sum_n \ell'_{kn} - \sum_n \ell_{kn}\right) \approx \frac{L}{\sqrt{I}}\sqrt{p'_k \times (1 - p'_k)} \equiv \mathfrak{s}'_k$$

$$\text{90\% CI of change} \approx \sum_n \ell'_{kn} - \sum_n \ell_{kn} \pm 1.645\mathfrak{s}'_k$$

Evaluate at $L = I$

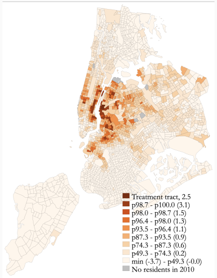
Application to Amazon's HQ2

Counterfactual: Amazon HQ2 in Long Island City

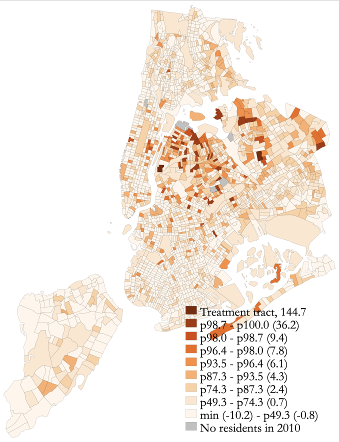
- Amazon's 2017 RFP for HQ2 with 50,000 employees elicited 238 proposals
- NYC proposed four possible sites (and controversial tax breaks)
- Split siting announced in 2018 would have put 25,000 employees in Long Island City
- Quantitative questions: What would happen to NYC neighborhoods with this local employment boom? Are these changes large relative to granular uncertainty?
- Granularity is important in bare-bones quantitative assessment (see Berkes and Gaetani 2020 for richer model)

Contrasting predictions for changes in residents

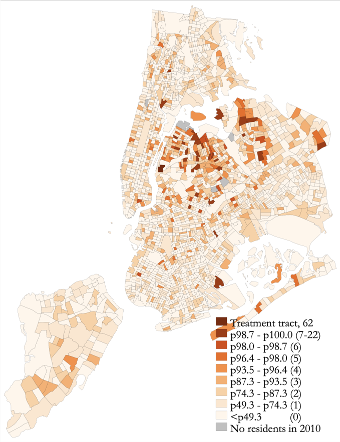
Calibrated-shares predictions are tightly tied to initial residents



Covariates-based model

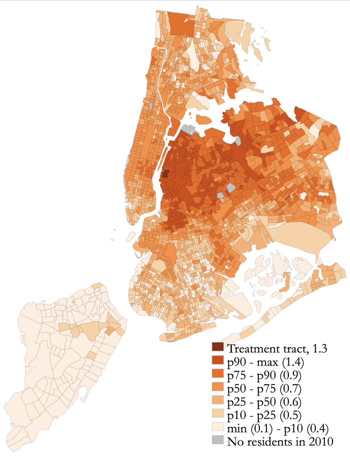


Calibrated-shares procedure

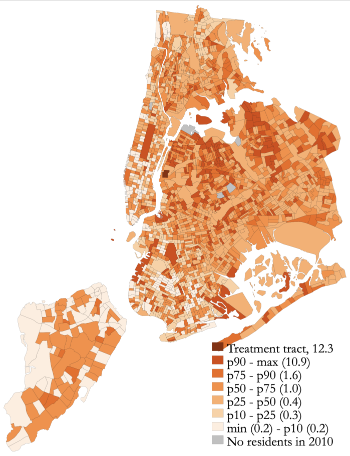


Residents working at treated tract

Contrasting predictions for changes in rents

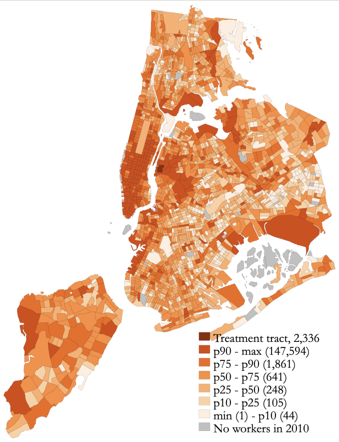


Covariates-based model

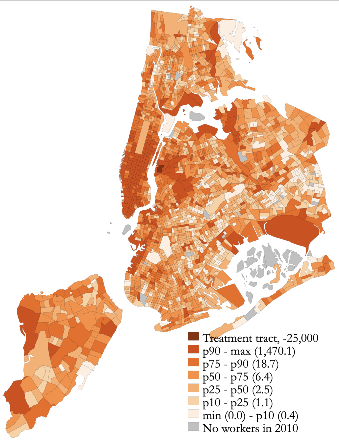


Calibrated-shares procedure

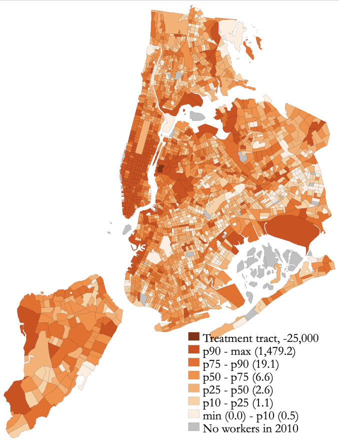
Predictions for changes in workers



Number of workers



Covariates-based model

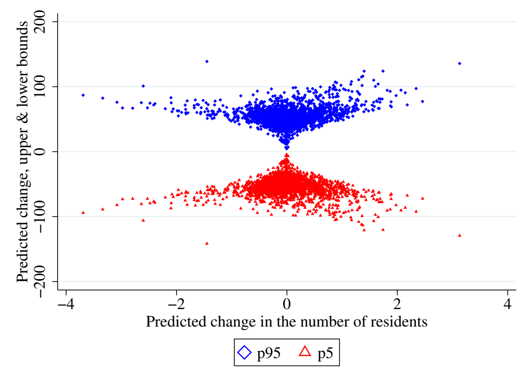


Calibrated-shares procedure

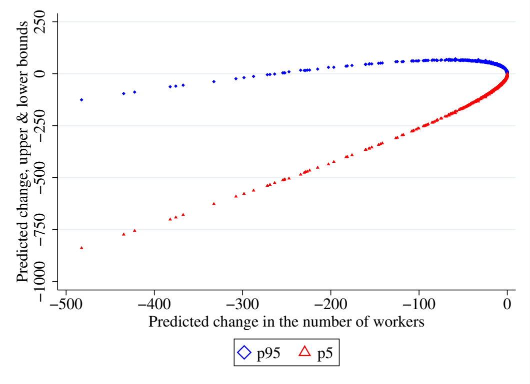
▸ Wage

Granular uncertainty is large relative to predicted changes

Change in residents



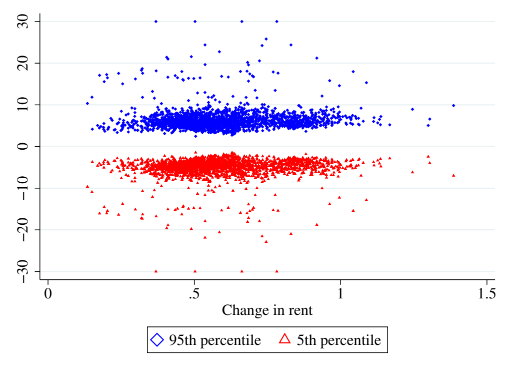
Change in workers



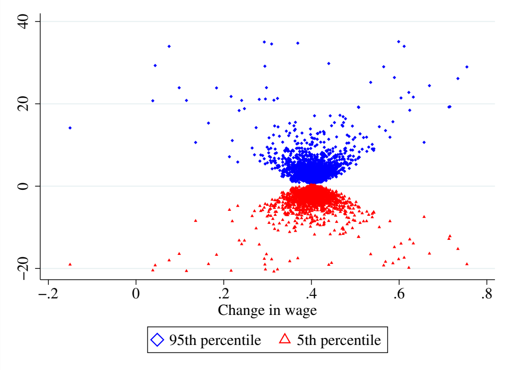
Residents: approx vs sim
Workers: approx vs sim
Calibrated shares

Granular uncertainty for predicted changes in prices

Predicted rent changes



Predicted wage changes



NOTES: The plots depict the 5th and 95th percentiles of predicted percentage-point change in price computed using the granular model. The horizontal axis displays the percentage-point change in mean price across 10,000 simulations. There are 43 tracts whose 5th percentile predicted wage change is greater than zero. The treated tract is excluded in wages panel.

Conclusions

Conclusions and next steps

- Finer spatial data are exciting but not a free lunch
- We need to evaluate the performance of applied GE models
- Monte Carlo and event studies: Calibrated-shares procedure performs poorly in granular empirical settings
- Researchers should use simulations to assess the finite-sample behavior of their counterfactual procedures
- Our granular model generates granular equilibrium outcomes and quantifies granular uncertainty accompanying counterfactual predictions
- Plain-vanilla logit assumption is simplest first step
- Our model is just as tractable, relies upon the same data, and coincides with the continuum case as $I \to \infty$

Thank you