ECON G6905

Topics in Trade

Jonathan Dingel

Spring 2025, Week 11

Columbia University in the City of New York

Today: Estimating and simulating discrete-choice models

My goals for today:

- Convey the connections between CES and logit in theory and estimation
- Introduce you to estimation and simulation of discrete-choice models in the context of international and urban economics
- Share tricks and resources that save lots of (human or computing) time

My papers using discrete-choice methods include

- How Segregated is Urban Consumption?
- Spatial Economics for Granular Settings

Blanket recommendations:

- Ken Train's *Discrete Choice Methods with Simulation* (2009, PDFs)
- Simulate your data-generating process to verify your procedures and code

Choosing location(s) is a discrete-choice problem

Long history of discrete-choice modeling in urban and transportation economics

- See McFadden's Nobel lecture; e.g., Peter Diamond & Robert Hall working on transport or multinomial logit vs gravity for BART forecasts
- The canonical nested-logit and GEV results are in McFadden (1978) "Modeling the Choice of Residential Location"
- Dennis Carlton on "The Location and Employment Choices of New Firms" in 1983
- Estimating logit model by Poisson regression is Guimarães, Figueiredo, Woodward "A Tractable Approach To The Firm Location Decision Problem" (2003)

Plain-vanilla logit case: individual i considers choice j (see Train 2009 Ch. 3)

- Utility $U_{ij} = V_{ij} + \epsilon_{ij}$
- Assume error is iid standard Gumbel (T1EV): $F\left(\epsilon_{ij}\right) = \exp(-\exp(-\epsilon_{ij}))$
- Choice probabilities are

$$\Pr\left(U_{ij} > U_{ij'}\ \forall j' \neq j\right) = \frac{\exp(V_{ij})}{\sum_{j'} \exp\left(V_{ij'}\right)}$$

Import-sourcing decisions are discrete-choice problems

Neoclassical trade models feature discrete-choice problems

- Selecting the least-cost provider of each good is at the heart of the model
- In the Eaton and Kortum (2002) formulation, this is probabilistic and a discrete-choice problem

Rewrite the plain-vanilla logit case from previous slide to be a cost-minimization problem with multiplicative error term

- Cost $\ln c_{ji} = \ln c_j + \ln \tau_{ji} - \epsilon_j$
- Least-cost probability

$$\Pr\bigl(\ln c_{ji} < \ln c_{j'i} \; \forall j' \neq j\bigr) = \Pr\bigl(-\ln c_{ji} > -\ln c_{j'i} \; \forall j' \neq j\bigr) = \frac{1/(c_j \tau_{ji})}{\sum_{j'} 1/(c_{j'} \tau_{j'i})}$$

Gumbel CDF is $F(\epsilon) = \exp(-\exp((\mu - \epsilon)/\beta))$. Standard Gumbel is $\mu = 0, \beta = 1$.
The Frechet distribution is the log-Gumbel distribution.

IIA logit demand implies CES market demand

Constant-elasticity demand functions are closely related. See Anderson, de Palma, Thisse (1992) book (and 1987 Economics Letters and 1988 IER)

Logit

$$\Pr\left(U_{ij} > U_{ij'}\ \forall j' \neq j\right) = \frac{\exp(V_{ij})}{\sum_{j'} \exp\left(V_{ij'}\right)}$$

let $V_{ij} = (1-\sigma)\ln\left(p_j \tau_{ji}\right) \implies$

$$= \frac{(p_j \tau_{ji})^{1-\sigma}}{P_i^{1-\sigma}}$$

CES

$$X_{ji} = \frac{(p_j \tau_{ji})^{1-\sigma}}{P_i^{1-\sigma}} X_i$$

(note subscript order)

How to arrange upper-tier preferences to align with aggregate expenditure X_i? If 2nd-stage quantity is $q^*_{ji} = X_i/p_{ji}$, then $V_{ij} = (1-\sigma)\ln\left(p_j \tau_{ji}/X_i\right)$ [see AdPT 1987]

(This explains the contrasting subscript orders of Ch 3 & 4 in 2014 *Handbook*)

Logit MLE coincides with Poisson regression (1/2)

Estimation of constant-elasticity functions is closely related (Guimaraes, Figueiredo, Woodward 2003). Let $p_{ij} \equiv \Pr\left(U_{ij} > U_{ij'} \; \forall j' \neq j\right) \propto \exp\left(\boldsymbol{\beta}'\mathbf{z}_{ij}\right)$.

Simple case: let $p_{ij} = p_j$

$$\ln \mathcal{L}^{\text{logit}} = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \ln p_{ij} = \sum_{j=1}^{J} n_j \ln p_j$$

Consider the Poisson count variable with value n_j: $\mathbb{E}\left(n_j\right) = \lambda_j = \exp\left(\alpha + \boldsymbol{\beta}'\mathbf{z}_j\right)$.

$$\ln \mathcal{L}^{\text{P}} = \sum_{j=1}^{J} \left(-\lambda_j + n_j \ln \lambda_j - \ln n_j!\right) = \sum_{j=1}^{J} \left[-\exp\left(\alpha + \boldsymbol{\beta}'\mathbf{z}_j\right) + n_j \left(\alpha + \boldsymbol{\beta}'\mathbf{z}_j\right) - \ln n_j!\right]$$

$$\frac{\partial \ln \mathcal{L}^{\text{P}}}{\partial \alpha} = \sum_{j=1}^{J} \left[n_j - \exp\left(\alpha + \boldsymbol{\beta}'\mathbf{z}_j\right)\right] = 0 \implies \exp(\alpha) = \frac{N}{\sum_{j=1}^{J} \exp\left(\boldsymbol{\beta}'\mathbf{z}_j\right)}$$

Substitute α back in to obtain the concentrated log likelihood...

Logit MLE coincides with Poisson regression (2/2)

Substitute α back in to obtain the concentrated log likelihood:

$$\begin{aligned}
\ln \mathcal{L}^{P_c} &= -N + N \ln N - \sum_{j=1}^{J} n_j \ln \left(\sum_{j=1}^{J} \exp\left(\boldsymbol{\beta}'\mathbf{z}_j\right) \right) + \sum_{j=1}^{J} n_j \boldsymbol{\beta}'\mathbf{z}_j - \sum_{j=1}^{J} \ln n_j! \\
&= \sum_{j=1}^{J} n_j \ln p_j - N + N \ln N - \sum_{j=1}^{J} \ln n_j! \\
&= \ln \mathcal{L}^{\text{logit}} + \text{constant}
\end{aligned}$$

The logit MLE β and the Poisson MLE β are the same. We can therefore use software implementations of PPML with high-dimensional fixed effects to estimate logit models.

"How Segregated is Urban Consumption?"

Two questions:

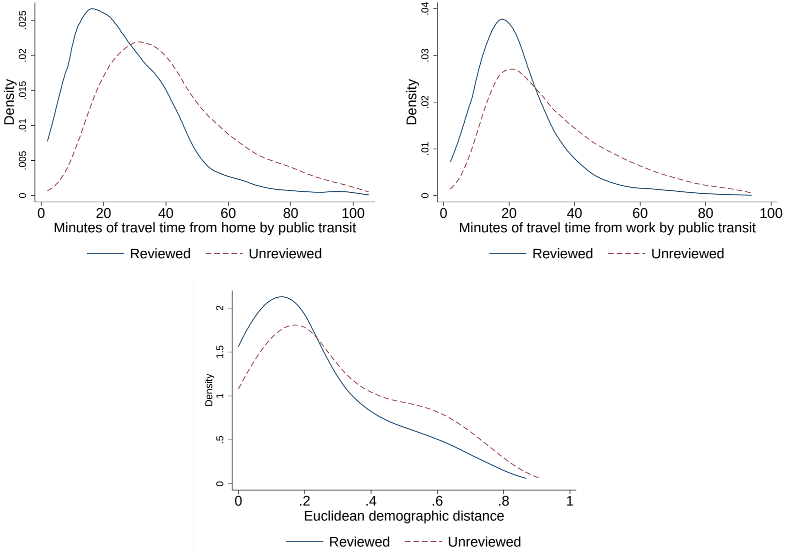
1. How segregated is urban consumption?
2. What are roles of spatial frictions, social frictions, and heterogeneous tastes?

How we answer them:

- Gather data on individuals' restaurant visits within New York City
- Infer spatial and social frictions from behavior by estimating a discrete-choice model of individuals' visit decisions
- Use model-predicted consumer behavior to measure consumption segregation

Here, I'll focus on the estimation and computational elements

Observed behavior, spatial frictions, social frictions



Demand estimation in “How Segregated is Urban Consumption?”

In a sense, this is “gravity within the city”, but the discrete-choice formulation is clearer in important respects

- Model of review-writing behavior necessitated by Yelp data
- Estimate parameters when data do not identify origin-mode pair l (e.g., from work by car)
- McFadden (1978) sampling exploits IIA to make computation feasible

Behavioral model

- Individual i decides at time t whether to visit venue j in choice set $\mathcal{J}$.
- Trip may originate from one of six locations l,
  $l \in \mathcal{L} = \{\text{car}, \text{public transit}\} \times \{\text{home, work, commute}\}$

$$U_{ijlt} = \underbrace{\gamma_l^1 X_{ijl}^1 + \gamma^2 X_{ij}^2 + \beta^1 Z_j^1 + \beta^2 Z_{ij}^2}_{=V_{ijl}} + \nu_{ijlt}$$

- Dummy $d_{ijlt} = 1$ if i visits j from l at t
- We observe dummy $d_{ijt} = \sum_{l \in \mathcal{L}} d_{ijlt}$
- Assume that ν_{ijlt} have type I extreme value distribution, independent across individuals, restaurants, time periods, and locations:

$$\Pr(d_{ijt} = 1 | X, Z, \mathcal{J}; \gamma, \beta) = \frac{\sum_{l \in \mathcal{L}} \exp(V_{ijl})}{\sum_{j' \in \mathcal{J}} \sum_{l \in \mathcal{L}} \exp\left(V_{ij'l}\right)}$$

Specification

$$V_{ijl} = \gamma_l^1 X_{ijl}^1 + \gamma^2 X_{ij}^2 + \beta^1 Z_j^1 + \beta^2 Z_{ij}^2$$

- Spatial frictions: X_{ijl}^1 is log minutes of transit time with mode-origin-specific disutilities γ_l^1
- Social frictions: X_{ij}^2 contains demographic differences between user i's home tract (or i's own identity) and the tract of restaurant j
- Restaurant characteristics: Z_j^1 contains venue rating, 4 price dummies, 9 cuisine dummies, venue tract median household income, and 28 area dummies
- User-restaurant characteristics: Z_{ij}^2 contains rating and price interacted with i's home tract median income, percentage and absolute percentage difference in median household incomes between i and j tracts

Identification: Reviews, not visits

We assume that

- users do not review unvisited venues
- users only review venues once
- probability of writing a review is independent of *ex ante* venue characteristics

$$\begin{aligned}\Pr\left(d^r_{ijt} = 1 | X, Z, \mathcal{J}; \gamma, \beta\right) &= \Pr\left(d^r_{ijt} = 1 | d_{ijt} = 1, \cdot; \cdot\right) \times \Pr(d_{ijt} = 1 | \cdot; \cdot) \\ &= w_{it} \times \mathbf{1}\{j \neq 0, j \notin D^r_{it}\} \times \Pr(d_{ijt} = 1 | \cdot; \cdot)\end{aligned}$$

If $\Pr\left(d^r_{ijt} = 1 | d_{ijt} = 1, \cdot; \cdot\right)$ depends on some restaurant characteristic in Z

- Coefficient on that characteristic would be biased
- Estimates of spatial and social frictions could still be asymptotically unbiased

Identification: Reviews, not visits

We assume that

- users do not review unvisited venues
- users only review venues once
- probability of writing a review is independent of *ex ante* venue characteristics

$$\begin{aligned}\Pr\left(d^r_{ijt} = 1 | X, Z, \mathcal{J}; \gamma, \beta\right) &= \Pr\left(d^r_{ijt} = 1 | d_{ijt} = 1, \cdot; \cdot\right) \times \Pr(d_{ijt} = 1 | \cdot; \cdot) \\ &= w_{it} \times \mathbf{1}\{j \neq 0, j \notin D^r_{it}\} \times \Pr(d_{ijt} = 1 | \cdot; \cdot)\end{aligned}$$

If $\Pr\left(d^r_{ijt} = 1 | d_{ijt} = 1, \cdot; \cdot\right)$ depends on some restaurant characteristic in Z

- Coefficient on that characteristic would be biased
- Estimates of spatial and social frictions could still be asymptotically unbiased

Computation: Large choice sets

Users choose amongst thousands of NYC restaurants

- McFadden (1978): estimate conditional logit model's parameters using a choice set S_{it} that is strict subset of actual choice set $\mathcal{J}$
- Construct S_{it} by including i's observed choice at period t plus a random subset of the other alternatives included in the set $J_i' = \mathcal{J} / \{D_i^r \cup \{j = 0\}\}$
- Select unchosen venues of S_{it} with equal probability from J_i'

$$\Pr\left(d_{ijt}^r = 1 | X, Z, S_{it}; \gamma, \beta\right) = \frac{\sum_{l \in \mathcal{L}} \exp\left(V_{ijlt}\right)}{\sum_{j' \in S_{it}} \sum_{l \in \mathcal{L}} \exp\left(V_{ij'lt}\right)}$$

$$L = \sum_i \sum_t \sum_{j \in S_{it}} \left\{ d_{ijt}^r \log\left(\Pr\left(d_{ijt}^r = 1 | X, Z, S_{it}; \gamma, \beta\right) \right) \right\}$$

- We construct S_{it} with 20 elements

Details of McFadden (1978) sampling

Denote by $\pi(S_{it}|d^*_{ijt} = 1, J'_{it})$ the probability of assigning the set S_{it} to an individual i who reviewed venue j at t. Our sampling scheme means

$$\pi(S_{it}|d^*_{ijt} = 1, J'_{it}) = \kappa_{it} \times \mathbf{1}\{j \in S_{it}\}$$

where $\kappa_{it} \in (0,1)$ is a constant determined by numbers of venues in S_{it} and J'_{it}. The resulting probability of reviewing restaurant j given sampling is

$$\Pr\left(d^*_{ijt} = 1|d^*_{it} = 1X, Z, \mathcal{J}; \gamma, \beta\right) = \mathbf{1}\{j \neq 0, j \notin D^r_{it}\} \times \Pr(d_{ijt} = 1|\cdot;\cdot)$$

$$\implies \Pr\left(d^*_{ijt} = 1|d^*_{it} = 1, X_i, Z_i, S_{it}; (\gamma, \beta)\right) = \frac{\mathbf{1}\{j \in S_{it}\} \sum_{l \in \mathcal{L}} \exp(V_{ijl})}{\sum_{j' \in S_{it}} \sum_{l \in \mathcal{L}} \exp\left(V_{ij'l}\right)}$$

is the probability that i reviews restaurant j at period t conditional on a randomly drawn set S_{it} and that i writes a review at t.

McFadden (1978) shows that maximizing this likelihood is a consistent estimator (obviously larger standard errors from not using all observations).

Embracing the independence of irrelevant alternatives

Logit specification features IIA across venues j and origins l

- Functional form makes the problem computationally feasible: McFadden (1978) sampling works because the ratio of probabilities for any two venues is the same whether or not other venues are available
- Estimation with randomly sampled S_{it} yields stable results, consistent with IIA (the other Hausman and McFadden (1984) test)
- See Train (2009) pages 49-50 on tests of IIA
- Can go to other extreme: IIA across venues j and identical across origins $l \rightarrow$ covariate is $X_{ij}^1 = \min_l X_{ijl}^1$
- This IIA at individual level, not aggregate market shares
- Plausibility of IIA depends on the demand system you model; rejection of IIA for country-level trade flows does not mean IIA is a bad description of individuals choosing restaurants with detailed characteristics

Simulating discrete-choice models

Some reasons to simulate the data-generating process.

- Verify that your numerical optimization code behaves correctly when the estimated model is correctly specified
- Construct confidence intervals for parameter estimates and downstream objects (relevant for finite samples or finite economies)
  - "Parametric bootstrap" of estimates (e.g., Davis, Dingel, Monras, Morales 2019)
  - Demonstrating finite-sample bias (e.g., critique of calibrated-shares procedure in Dingel Tintelnot 2023)
  - Building confidence intervals for counterfactual outcomes (e.g., Section 5 of Dingel Tintelnot 2023)

(A distinct role for simulation is in estimating models that do not have closed-form expressions for choice probabilities. For example, multinomial probit and, more importantly, mixed (random-coefficients) logit. See Train (2009) starting with Chapter 5.)

Simulating random variables

- The CDF of a random variable has standard uniform distribution $U(0,1)$
- Any machine can simulate draws of $u \sim U(0,1)$, so invertible CDF lets you simulate that random variable.
- Standard Gumbel: $F(\epsilon) = \exp(-\exp(-\epsilon)) \implies \epsilon = -\ln(-\ln(u))$
- Drawing 100 million standard Gumbel realizations in Julia, Stata, and R

```
idiosyncratic = -log.(-log.(rand(100_000_000)))
```

```
set obs 100000000
gen idiosyncratic = -ln(-ln(runiform()))
```

```
idiosyncratic = -log(-log(runif(100000000)))
```

Simulate what? Simulating idiosyncratic preferences vs choices

Consider two ways to simulate draws when using your estimated model as the data-generating process (e.g., $U_{ij} = V_{ij} + \epsilon_{ij}$ and $F(\epsilon_{ij}) = \exp(-\exp(-\epsilon_{ij}))$)

1. Simulate the logit error terms ϵ_{ij} and compute utility-maximizing choices
2. Simulate the outcomes given by $\Pr\left(U_{ij} > U_{ij'}\ \forall j' \neq j\right) = \frac{\exp(V_{ij})}{\sum_{j'} \exp\left(V_{ij'}\right)}$

While logically equivalent, these differ tremendously in computational burden

- Time (in seconds) to produce 100 million Gumbel draws:

  $\sim$ 3s in Julia, $\sim$ 5s in Stata, $\sim$ 3.5s in R

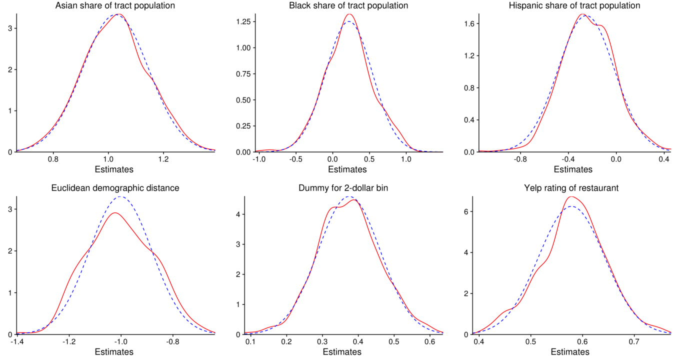
- Taking 100 draws from a multinomial distribution with 1 million possible choices takes < 0.1 seconds

```
using Distributions; mean_util = rand(1_000_000);
@time choices = rand(Multinomial(100,exp.(mean_util) ./ sum(exp.(mean_util)
    )));
```

Parametric bootstrap for confidence intervals in DDMM

- Draw 500 samples from estimated model (same size as estimation sample)
- Estimate the model on each generated sample
- Distributions for social frictions and restaurant characteristics look like asymptotic normal distribution

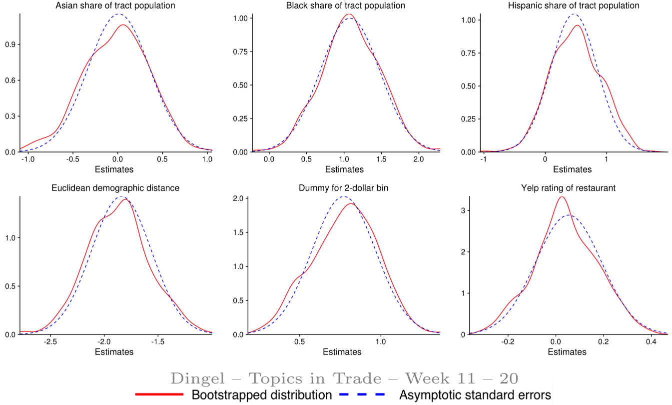
Asian reviewers: Social frictions



Parametric bootstrap for confidence intervals in DDMM

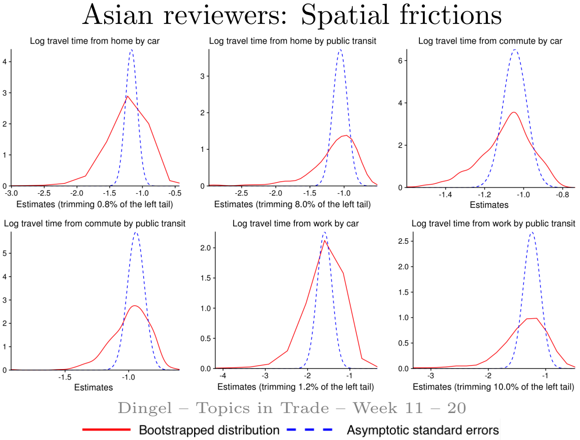
- Draw 500 samples from estimated model (same size as estimation sample)
- Estimate the model on each generated sample
- Distributions for social frictions and restaurant characteristics look like asymptotic normal distribution

Black reviewers: Social frictions



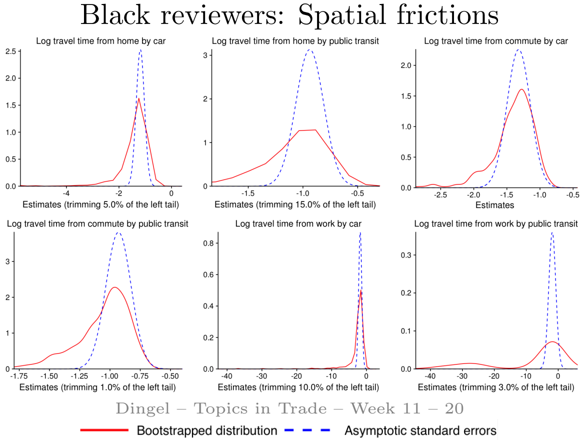
Parametric bootstrap for confidence intervals in DDMM

- Draw 500 samples from estimated model (same size as estimation sample)
- Estimate the model on each generated sample
- Distributions for spatial frictions have fat tails because of collinearity of same-origin modes (see Goldberger on multicollinearity and micronumerosity)



Parametric bootstrap for confidence intervals in DDMM

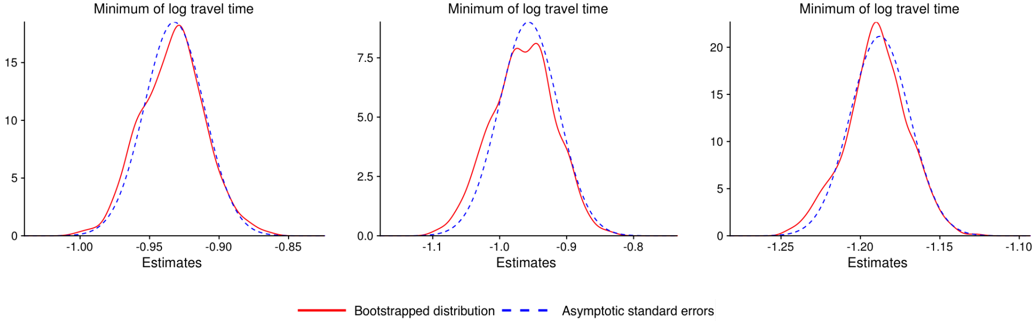
- Draw 500 samples from estimated model (same size as estimation sample)
- Estimate the model on each generated sample
- Distributions for spatial frictions have fat tails because of collinearity of same-origin modes (see Goldberger on multicollinearity and micronumerosity)



Parametric bootstrap for confidence intervals in DDMM

- Draw 500 samples from estimated model (same size as estimation sample)
- Estimate the model on each generated sample
- Distributions for spatial frictions have fat tails because of collinearity of same-origin modes (see Goldberger on multicollinearity and micronumerosity)

Spatial frictions in minimum-time specification $\nu_{ijlt} = \nu_{ijt}\ \forall l$



Parametric bootstrap of dissimilarity indices in DDMM

You can bootstrap any function of the parameters (i.e., any downstream object)

TABLE 6
Residential and Consumption Segregation

	Residential Dissimilarity (1)	Consumption Dissimilarity: Estimated (2)	No Spatial (3)	No Social (4)	Neither Friction (5)
	A. Dissimilarity Index				
Asian	.521	.315	.290	.245	.232
		[.305, .335]	[.280, .314]	[.233, .268]	[.222, .259]
Black	.653	.352	.322	.273	.260
		[.337, .397]	[.307, .372]	[.258, .320]	[.248, .309]
Hispanic	.486	.142	.114	.106	.088
		[.134, .162]	[.108, .137]	[.099, .125]	[.083, .109]
White	.636	.190	.153	.112	.093
		[.180, .209]	[.143, .174]	[.106, .130]	[.090, .112]
White or Hispanic	.470	.205	.189	.150	.156
		[.197, .236]	[.182, .224]	[.143, .182]	[.149, .191]

Note.—This table reports dissimilarity indices. Panel A reports the index for each demographic group's residential/consumption locations compared to members of all other demographic groups. Panel B reports the index for residential/consumption locations between each pair of demographic groups. The demographic group "other" is included in computations but not reported. Col. 1 reports indices based on tracts' residential populations. The remaining columns report venue-level dissimilarity indices based on the coefficient estimates in cols. 4–6 of table 2. Col. 2 uses the estimated coefficients. Col. 3 sets the coefficients on travel time covariates to zero. Col. 4 sets the coefficients on demographic-difference covariates to zero. Col. 5 sets the coefficients on travel time and demographic difference covariates to zero. Bootstrapped 95 percent confidence intervals from 496 draws are reported in brackets.

Monte Carlo simulations to show finite-sample bias

Recall Dingel & Tintelnot (2023) critique of calibrated-shares procedure

- What parameterization of the baseline equilibrium should be used to compute counterfactual outcomes?
- Calibrated-shares procedure employs observed shares
- Covariates-based approach fits more parsimonious spec of commuting costs δ_{kn}
- An excessively flexible parameterization risks overfitting idiosyncratic noise
- D & T use Monte Carlo simulations to demonstrate the overfitting problem

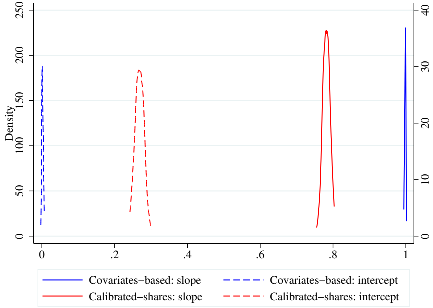
Monte Carlo in DT (2023): Applying each procedure to finite data

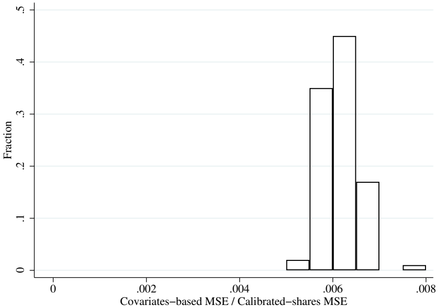
- DGP is estimated covariates-based model for NYC in 2010
- Simulated "event": ↑ productivity of 200 Fifth Ave tract by 9%
- Apply calibrated-shares procedure and covariates-based approach
  (Increase A_n to match total employment increase in simulated data)
- Does the procedure predict the change in the number of commuters from each residential tract working in the "treated" tract?
- Regress "true" changes on predicted changes (2160 obs per simulation)
  Ideally, want slope = 1 and intercept = 0
- Compute forecast errors (MSE for "true" vs predicted changes)

Monte Carlo in DT (2023): Calibrated-shares procedure overfits

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

Changes in commuter counts ($\ell'_{k\bar{n}} - \ell_{k\bar{n}}$)



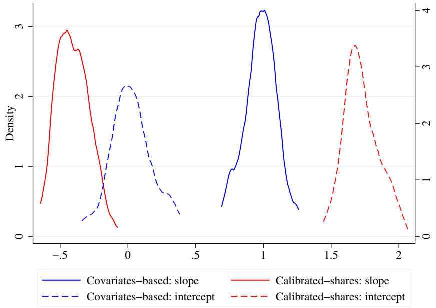


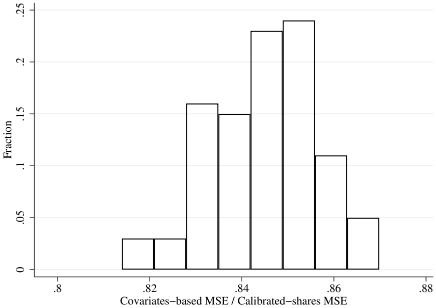
I	2.5	5	12.5	25	50	125	250	2560
Calibrated-shares: slope	0.782	0.876	0.948	0.974	0.986	0.995	0.997	1.000
Calibrated-shares: intercept	0.269	0.153	0.064	0.032	0.017	0.007	0.004	0.000
Calibrated-shares: MSE	0.225	0.113	0.045	0.023	0.011	0.005	0.002	0.000

Monte Carlo in DT (2023): Calibrated-shares procedure overfits

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

Changes in commuter counts $(\ell'_{k\bar{n}} - \ell_{k\bar{n}})$ via finite-sample draws from pre- and post- DGPs





I	2.5	5	12.5	25	50	125	250	2560
Calibrated-shares: slope	-0.408	0.194	0.669	0.835	0.913	0.968	0.982	0.998
Calibrated-shares: intercept	1.724	0.982	0.404	0.202	0.106	0.040	0.022	0.002
Calibrated-shares: MSE	17.022	8.486	3.400	1.699	0.851	0.340	0.169	0.017

A spatial model with a finite number of individuals

Goal: examine the sensitivity of counterfactual outcomes to the idiosyncratic component of individual decisions

In the limit ($I \to \infty$), the equilibrium of our model with an integer number of individuals is (almost surely) the equilibrium of the continuum model

Modeling concerns raised by the integer number of individuals:

- 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 Υ

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 $\Upsilon \equiv (L, \{A_n\}, \{T_k\}, \{\bar{\delta}_{kn}\}, \{\lambda_{kn}\}, \alpha, \epsilon, \sigma)$ and have (common) point-mass beliefs $\tilde{r}_k$ and $\tilde{w}_n$ about land prices and wages
- Worker i knows own idiosyncratic preferences $\{\nu^i_{kn}\}$ but not the full set of idiosyncratic residence-workplace draws $\boldsymbol{\nu}^I$

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 $\Upsilon \equiv (L, \{A_n\}, \{T_k\}, \{\bar{\delta}_{kn}\}, \{\lambda_{kn}\}, \alpha, \epsilon, \sigma)$ and have (common) point-mass beliefs $\tilde{r}_k$ and $\tilde{w}_n$ about land prices and wages
- Worker i knows own idiosyncratic preferences $\{\nu^i_{kn}\}$ but not the full set of idiosyncratic residence-workplace draws $\boldsymbol{\nu}^I$

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 $\{\ell_{kn}\}$ and economic primitives Υ, **a trade equilibrium** is a set of wages $\{w_n\}$ and land prices $\{r_k\}$ that clears all markets.

Commuting equilibrium with a finite number of individuals

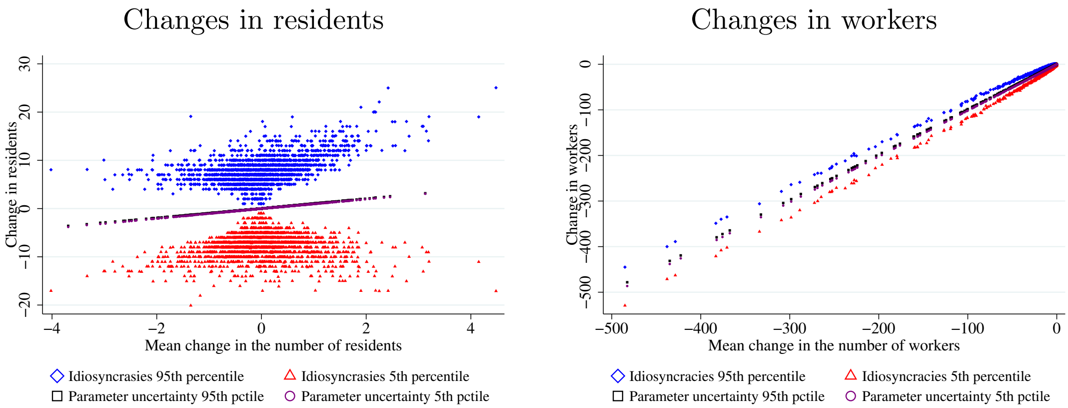
Given primitives Υ, idiosyncratic residence-workplace draws $\boldsymbol{\nu}^I$, and point-mass beliefs $\{\tilde{w}_n\}, \{\tilde{r}_k\}$, a **commuting equilibrium with a finite number of individuals**, I, is defined as a labor allocation $\{\ell_{kn}\}$, wages $\{w_n\}$, and land prices $\{r_k\}$ such that

- $\ell_{kn} = \frac{L}{I} \sum_{i=1}^{I} \mathbf{1}\{\tilde{U}^i_{kn}(\boldsymbol{\nu}^I) > \tilde{U}^i_{k'n'}(\boldsymbol{\nu}^I)\ \forall (k', n') \neq (k, n)\}$; and
- wages $\{w_n\}$ and land prices $\{r_k\}$ are *a trade equilibrium* given the labor allocation $\{\ell_{kn}\}$.

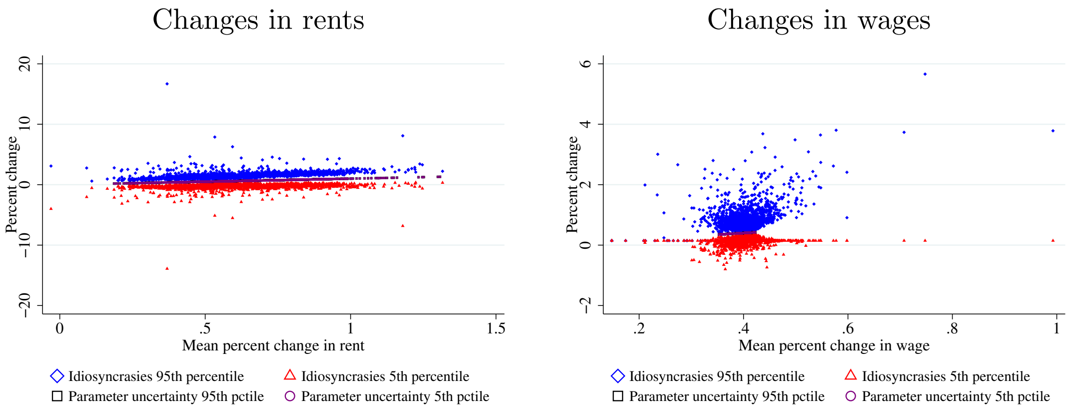
Comparison with continuum model

- Idiosyncratic residence-workplace draws $\boldsymbol{\nu}^I \rightarrow$ distributions of equilibrium quantities and prices (for given primitives Υ)
- Mean equilibrium outcomes:
  - Mean commuter counts coincide with those from the continuum model $\frac{\ell_{kn}}{L} = \mathbb{E}\left[\Pr\left(U^i_{kn} > U^i_{k'n'}\ \forall (k',n') \neq (k,n)\right)\right]$
  - Land prices and wages are solved from a non-linear system of equations
- Variance of equilibrium outcomes due to idiosyncrasies
  - Confidence interval for residents, workers, wages, and prices
- In counterfactual exercises: Change from Υ to Υ' for given $\boldsymbol{\nu}^I$
- The set of individuals who change their decisions in response to the change in economic primitives depends on the particular realized vectors of idiosyncratic preferences
- The dispersion in this distribution represents uncertainty about counterfactual predictions stemming from individual idiosyncrasies

Sizable uncertainty about predicted changes from idiosyncrasies



Sizable uncertainty about predicted changes from idiosyncrasies



Simulate what? Simulating idiosyncratic preferences vs choices

Holding $\boldsymbol{\nu}^I$ fixed requires simulating 8 trillion Gumbel draws

- Change from Υ to Υ' for given $\boldsymbol{\nu}^I$: ~ 100 CPU hours per simulation
- Contrast outcome distributions for Υ and Υ': < 1 CPU hour per simulation

Do the cheaper simulations first

- Can you rule out uncertainty attributable to individual idiosyncrasies? Repeatedly draw multinomial realizations to compute a distribution of differences between finite-sample outcomes at counterfactual parameters Υ' and the expected outcome at baseline parameters Υ.
- This distribution of differences is more dispersed than the distribution of counterfactual changes from the model with a finite number of individuals that fixes the vector of idiosyncratic preferences $\boldsymbol{\nu}^I$.

Summary

- CES and logit are close cousins; constant-elasticity siblings, really
- Reading a couple chapters of Train (2009) goes a long way
- Smart choices can speed your computations by orders of magnitude
- Econometrics by simulation is often a good starting point

Next week: Spatial environmental economics