E-commerce and Regional Inequality:

A Trade Framework and Evidence from Amazon's Expansion

Elmer Zongyang Li

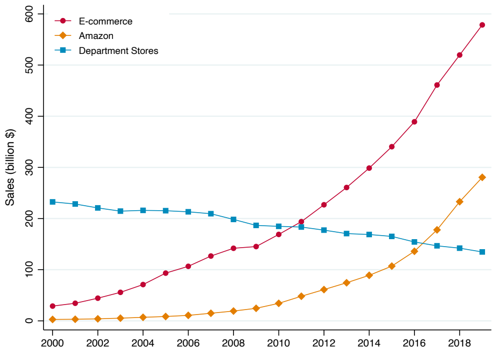
International Monetary Fund

January, 2025

ASSA Meetings

Disclaimer: The views expressed herein are those of the author and should not be attributed to the IMF, its Executive Board, or its management

Brick-and-motor vs. E-commerce



- Secular ↑ online retail sales (**e-commerce**)
- "Opening to trade" challenges *regional equality*
  - Comparative advantages, worker specializations, input-output linkages
  - The unique nature of online retailing may exacerbate

This Paper

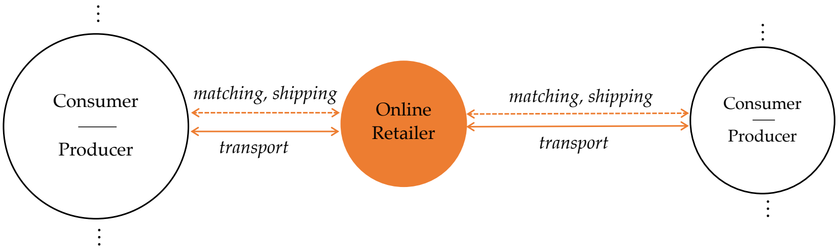
E-commerce as a unique trade shock $\Rightarrow$ Spatial **GE** and **reallocation** (welfare, empl. dispersion)

- **Empirics**: New facts on Amazon sales, retailers, facilities
  - Online retailer spatial concentration, sales & trade
- **Theory**: multi-region & -sector spatial (retail) trade model
  - Consumer search & shipping
  - Location choice of online retailer $\Rightarrow$ $\uparrow$spatial concentration
- **Policy**: place-based public finances

Contribution: new data & extend spatial trade theory $\Rightarrow$ e-commerce

Data Sources

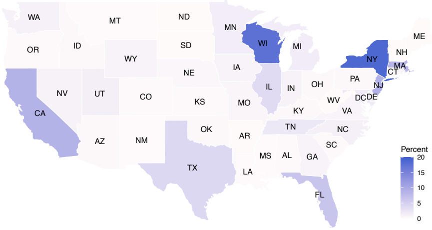
- Amazon Retailers and Products (Keepa.com)
  - Universe of products on Amazon (36 categories, 2016-2020, 0.1%)
  - Information on prices, sales, sellers' addresses, FBA status
- Amazon Facilities (MWPVL)
  - Addresses, square feet, date, type. Focus on large & general ones
- DOT Commodity Flow Survey (CFS)
- Other: CBP, BEA, ACS, Geography (topography, climate)



Empirical Pattern - Spatial Concentration

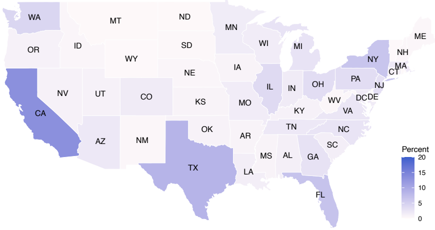
(1) Regional Share of E-commerce Sales

(HHI = 0.16)



(2) Regional Share of Retail Sector Sales

(HHI = 0.05)



- 1a: Online retail sales are more concentrated than average retail sales...
- 1b: ...and those that are FBA more concentrated than non-FBA details
- 2: Durable/standardized ones are less concentrated details
- 3a: Online retail is less correlated with population or taxes details
- 3b: ...and the concentration aligns with truck routes details

Empirical Patterns - Interregional Trade

- 4a: Origins with ↑ online retailers export more tradable goods
- 4b: Destinations with ↑ online retailers import less tradable goods
- 5: Regions near to fulfillment centers have more trade flows

Dependent Variable:	$\Delta \ln$ (Shipment)	
Δ share (%) of online sellers - origin	3.5***	
	[0.8]	
Δ share (%) of online sellers - destination	-1.4*	
	[0.7]	
$\Delta \ln$ (bilateral distance via Amazon facility)		4.92*
		[2.53]
Origin, destination FE	✓	
Industry FE	✓	
Observations	24,693	24,693
R-squared	0.20	0.19

Model Summary details

- N regions; $2+2$ sectors: (home, service) & (dur, non-dur)
- 3 subsectors: M (manufacturer), R (online retailer), B (brick-and-mortar)

1. **Demand**: Sequential directed search → CES w/. demand shifter

$$C_n^j = \left[(c_{nn}^B)^{\frac{\sigma^j-1}{\sigma^j}} + \mu \sum_{m=1}^{N} \int_0^{O_m^j} (c_{nm}^R(i))^{\frac{\sigma^j-1}{\sigma^j}} di \right]^{\frac{\sigma^j}{\sigma^j-1}}$$

2. **Intermediate**: Ricardian (EK) → manuf. trade flow

3. **Online Seller**: Location choice → concentration, retail trade flow

$$m^* = \arg\min_m \left\{ \sum_n \left(\tilde{\sigma} \frac{c_m^{j,R}}{z_m^{j,R}} \frac{\kappa_{nm}^R}{P_n^{j,R}} \right)^{\sigma^j-1} \cdot \frac{1}{\eta^j X_n} \right\} \quad \left(\equiv \frac{\tilde{\sigma}\xi_m^j}{z_m^{j,R}} \right)$$

$$\Psi_m^j = P(m = m^* \cap c_m^{j,R} < \bar{c}_m^j); \; x_{nm}^{j,R} = \frac{\Psi_m^j (\kappa_{nm}^R c_m^{j,R}/\mu)^{1-\sigma}}{\sum_h \Psi_h^j (\kappa_{nh}^R c_h^{j,R}/\mu)^{1-\sigma} + \frac{1}{O^j} (c_n^{j,B})^{1-\sigma}}$$

4. **Worker**: Roy labor supply

E-commerce and Calibration

- Welfare: real income per capita $W_n = \frac{Y_n/L_n}{P_n}$, its change:

$$\hat{W}_n = \overbrace{\underbrace{\hat{w}_n^0 (\hat{\pi}_n^0)^{\frac{-1}{\nu_n}}}_{\text{non-emp. worker special.}}}^{\text{income effect}} \times \overbrace{\underbrace{\Pi_{j=1}^J (\hat{x}_{nn}^{j,B})^{\frac{-\eta_j}{\sigma^j - 1}}}_{\text{industry composition}} \underbrace{(\hat{c}_n^{j,R/B})}_{\text{input-output local pref.}}}^{\text{price effect}}$$

- External Calibration (2007) details
  - Fix w/. data or literature. Match untargeted sectoral incomes
- E-commerce (Δ 2007-2017) details
  - ↑ Match efficiency μ: 1.27 [1.46] (Dinerstein et. al 2018; Goldmanis et. al 2010)
  - ↓ Bilateral frictions $\hat{\kappa}_{ni}^R$: 0.97 [0.15] (Houde, Newberry & Seim 2021)
  - ↑ Online retailer spatial concentration Ψ_m^j (Keepa, targeted)

Welfare – Total

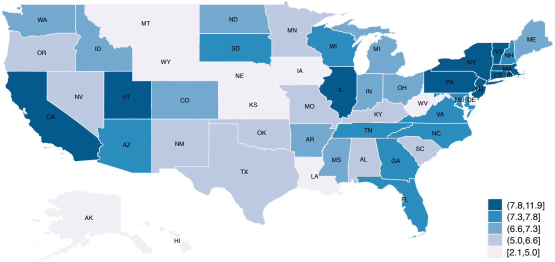


Figure: Total Welfare Change

- ↑ **welfare** overall (avg: 6.7 %)
  - States on the East and West Coasts experience larger welfare gains
  - Midwestern states see smaller increases

Welfare – Decomposition

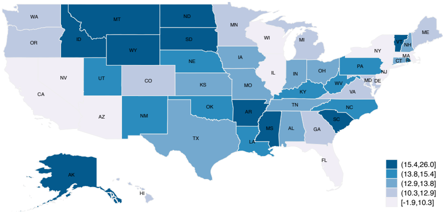


Figure: Price effects

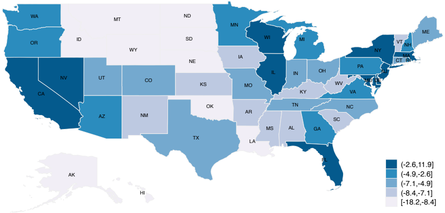


Figure: Income effects

- **Price effects** $\uparrow$ welfare (13.1%); **Income effects** $\downarrow$ welfare (5.4%)
  - States w/. CA in e-commerce and diverse industries (NY, MA, WI, CA, FL): Positive income effects due to $\uparrow$ online sales, wages
  - Midwestern: Negative income effects from competition and labor shifts. Lower initial online spending $\rightarrow$ Positive price effects

Result – Employment

Sector	All States Mean	All States Std. Dev.	Below 50th Percentile Online Sales Density Mean	Below 50th Percentile Online Sales Density Std. Dev.
Manufacturing	-4.3	(7.6)	-1.8	(1.1)
Online Retail	109.8	(97.8)	63.3	(64.8)
Brick-and-Mortar	-11.1	(8.0)	-8.6	(1.2)
Service	-1.6	(7.9)	1.2	(1.2)
Non-Employment	-1.3	(8.1)	1.7	(0.8)

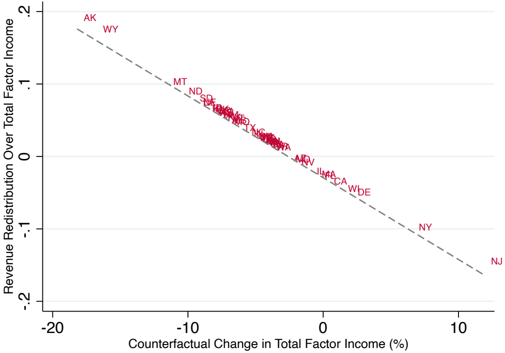
Table: Employment Changes by Sector and State Groups

- Reallocate from manufacturing/brick-and-mortar to online retail; non-employment $\downarrow$ by 0.5 ppts.
- Midwestern states shift more to service/non-employment sectors
- $\uparrow$ **inequality**: Gini 0.11→0.38

(Simple) Revenue Redistribution

- Government Objectives
  - Common welfare changes ($\forall n, \hat{W}_n = \frac{\hat{Y}_n}{\hat{P}_n} = k$), by manipulating $Y'_n \rightarrow \tilde{Y}'_n$
  - Same total surplus $\sum_{n=1}^{50}(\tilde{Y}'_n - Y_n) = B = \sum_{n=1}^{50}(Y'_n - Y_n)$

$$\Rightarrow \quad k = \frac{B+\sum_{n=1}^{50} Y_n}{\sum_{n=1}^{50} Y_n \cdot \frac{\tilde{P}_n}{P_n}} = 0.97; \text{ redistrib. amt} = (\tilde{Y}'_n - Y'_n) = Y_n k \frac{\tilde{P}_n}{P_n} - Y'_n$$



Conclusion

- E-commerce as unique trade shock
- **New facts** on online retailer spatial concentration (sellers, trade flows)
- **Spatial retail trade model** w/. location choices (search efficiency, elastic labor)
- Amazon $\Rightarrow$ efficiency equality tradeoff on welfare, empl.
  - $\downarrow$ prices, $\uparrow$ variety, but $\downarrow$ income and empl. adjmnt in Midwestern
  - Need national level revenue redistribution

Weitzman (1979) back

Search is *ordered*: Weitzman (1979) optimal stopping

- Assign thresholds/scores $\bar{v}_i$ st. $E[\max\{\hat{x}_i + \tilde{\epsilon}_i - \bar{v}_i, 0\}] = 0$, where $\hat{x}_i = \ln y - \ln p_i$
- Therefore, $\bar{v}_i = \hat{x}_i + \gamma_{\epsilon_i}^{-1}(\ln s_i)$, where $\gamma_{\epsilon_i}(z) = E[\max\{\epsilon_i - z, 0\}]$, decreasing function
- Search in decreasing order of the scores
- Stop if find a $\bar{v}_i$ exceeding all remaining

Proof of OSM to DCM back

Proposition: For any OSM, there is a DCM with same demand & payoff.

- Under OSM, consumer's optimal choice is the one for which $v_i^* = \min\{v_i, \bar{v}_i\}$ is largest (Armstrong and Vickers (2015), Armstrong(2017), Choi, Dai and Kim(2018)), where $\bar{v}_i = \hat{x}_i + \gamma_{\epsilon_i}^{-1}(\ln \mu_i) = \hat{x}_i + r(\ln \mu_i)$, and $\gamma_{\epsilon_i}(z) = E[\max\{\epsilon_i - z, 0\}]$, the upside gain function
- Consumer's demand for i, D_i is thus:
$$P[v_i^* > \max_{j \neq i} v_j^*] = \int_{-\infty}^{\infty} P[z > \max_{j \neq i} v_j^*] f_{v_i^*}(z; x_i, \hat{x}_i) dz = \int_{-\infty}^{\infty} \Pi_{j \neq i} F_{v_j^*}(z; x_j, \hat{x}_j) f_{v_i^*}(z; x_i, \hat{x}_i) d$$
- Under advertised price, $x_j = \hat{x}_j$, $\forall j$. D_i then simplifies to
$$\int_{-\infty}^{\infty} \Pi_{j \neq i} F_{\omega_j}(\epsilon_j) f_{\omega_i}(\epsilon_i) d\epsilon, \text{ where } \omega_i = \min\{\epsilon_i, r(\ln \mu_i)\}.$$
Thus, D_i is equivalent to the demand of a DCM: $v_i = x_i + \epsilon_i^{DC}$, iff

Proof of DCM to CES back

Proposition: The CES demand is a special case of DCM with extreme type I error.

The following proof follows Anderson, De Palma, and Thisse (1987, 1989) closely

- Consumer's utility $u_i = \ln c_i$, income y. Let price of i: $\tilde{p}_i = \mu_i p_i$
- Random utility/match value ϵ_i with i, st. net value:

  $v_i = \ln y - \ln \tilde{p}_i + \epsilon_i^{DC}$

  Further, re-scale $\epsilon_i^{DC} = \chi \tilde{\epsilon}_i$ st. $\tilde{\epsilon}_i$ mean 0 and unit variance
- The demand for i, D_i is then

$$P[v_i > \max_{j \neq i} v_j] = \int_{-\infty}^{\infty} \Pi_{j \neq i} F_{\epsilon_j^{DC}}(\epsilon_j^{DC}) f_{\epsilon_i^{DC}}(\epsilon_i^{DC}) d\epsilon.$$

- And if $\tilde{\epsilon}_i$ is distributed extreme type I, D_i then simplifies to

Market Clearing Conditions back

- Retail and intermediate goods:

$$X_n^{R,j} = \sum_{i=1}^{N} x_{in}^{R,j} (I_i L_i), \text{where } I_i L_i = \sum_{k=0}^{J} [r_i^{g,k} g_i^{R,k} + \sum_{K=M,R} (r_i^{h,k} h_i^{K,k} + w_i^k l_i^{K,k})] - \Omega_i,$$

$$X_n^{M,j} = \sum_{i=1}^{N} (1 - \gamma_i^j) x_{in}^{M,j} X_i^{R,j}.$$

- Trade balance:

$$\sum_{j=0}^{J} \sum_{i=1}^{N} (x_{ni}^{M,j} X_n^{M,j} + x_{ni}^{R,j} X_n^{R,j}) + \Omega_n = \sum_{j=0}^{J} \sum_{i=1}^{N} (x_{in}^{M,j} X_i^{M,j} + x_{in}^{R,j} X_i^{R,j}).$$

- Labor market: $w_n^{M,j} l_n^{M,j} = \beta_n X_n^{M,j}, \;\; w_n^{R,j} l_n^{R,j} = \gamma_n^j m_n^{R,j} \beta_n X_n^{R,j}$
- Structure: $r_n^h h_n^{M,j} = (1 - \beta_n) X_n^{M,j}, \;\; r_n^h h_n^{R,j} = \gamma_n^j \frac{1}{\rho_n^{R,j}} (1 - \beta_n) X_n^{R,j}$
- Capital: $r_n^g g_n^{R,j} = (\frac{\rho_n^j - 1}{1 - \beta_n}) w_n^{R,j} \pi_n^{R,j} L_n$

Comparative Statics back

- Employment shares:

$$\hat{\pi}_n^0 = \frac{\hat{A}_n^0 (\hat{w}_n^0)^{\nu_n}}{\hat{\Phi}_n}, \quad \hat{\pi}_n^{K,j} = \frac{\hat{A}_n^{K,j} (\hat{w}_n^{K,j})^{\nu_n}}{\hat{\Phi}_n}, \text{ where } \hat{\Phi}_n = \sum_{h=0}^{J} \sum_{K=M,R} \pi_n^{K,h} \hat{A}_n^{K,h} (\hat{w}_n^{K,h})^{\nu_n}.$$

- Input costs: $\hat{c}_n^{M,j} = \hat{\omega}_n^{M,j}, \quad \hat{c}_n^{R,j} = (\hat{\rho}_n^{R,j} \hat{\omega}_n^{R,j})^{\gamma_n^j} (\hat{P}_n^{M,j})^{1-\gamma_n^j}$, where

$$\hat{\omega}_n^{K,j} = \hat{w}_n^{K,j} (\hat{l}_n^{K,j})^{\beta_n} = (\hat{w}_n^{K,j})^{1+\beta_n} (\hat{\pi}_n^{K,j})^{\frac{(\nu_n - 1)\beta_n}{\nu_n}}, \text{and } \hat{P}_n^{M,j} = \Big(\sum_{i=1}^{N} x_{ni}^{M,j} (\hat{\kappa}_{ni}^{M} \hat{c}_i^{M,j})^{-\theta^j} \hat{T}_i^j\Big)^{\frac{-1}{\theta^j}}.$$

- Trade shares: $x_{ni}^{\prime M,j} = x_{ni}^{M,j} \left(\frac{\hat{\kappa}_{ni}^{M} \hat{c}_i^{M,j}}{\hat{P}_n^{R,j}}\right)^{-\theta_j} \hat{T}_i^j, \quad x_{ni}^{\prime R,j} = x_{ni}^{R,j} \left(\frac{\hat{\kappa}_{ni}^{R} \hat{c}_i^{R,j}}{\hat{\mu}_{ni}^j \hat{P}_n^{R,j}}\right)^{1-\sigma^j}.$
- Market clearing:

$$X_n^{\prime R,j} = \textstyle\sum_{i=1}^N x_{in}^{\prime R,j} \eta^j \left[\sum_{k=0}^J \left(\frac{1}{1-\beta_i}\right) \left(\hat{\rho}_i^{R,k} \hat{w}_i^{R,k} \hat{l}_i^{R,k} \rho_i^{R,k} w_i^{R,k} L_i^{R,k} + \hat{w}_i^{M,k} \hat{l}_i^{M,k} w_i^{M,k} L_i^{M,k}\right) - \Omega_i\right],$$

$$X_n^{\prime M,j} = \sum_{i=1}^{N} (1 - \gamma_i^j) x_{ni}^{\prime M,j} X_n^{\prime R,j},$$

Aggregate Trade back

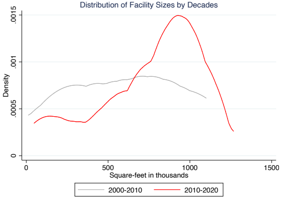
- Pareto productivity: $P(Z^j < z) = G^j(z) = 1 - z^{-\rho}$
- Enter: $\sum_n \left(\frac{p^j_{nm/\mu}}{p^R_{nj}}\right)^{1-\sigma} \eta^j_n \geq \omega^j_m f^j_m$. $c^j_m = \mu \left(\frac{\sigma}{\tilde{\sigma}_n}\right)^{1-\sigma} \left[\frac{w^j_m f^j_m}{\sum_n (k^R_{nm}/p^R_{nj})^{1-\sigma} \frac{1}{y_n}}\right]^{\frac{1}{1-\sigma}}$
- Bilateral trade shares

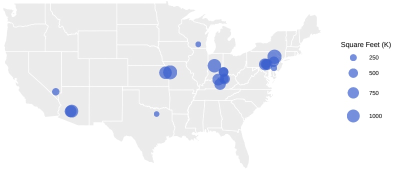
$$x^{j,R}_{nm} = \frac{\lambda Y_m \left(\left(w^{j,R}_m\right)^{\gamma^j} \left(P^{j,M}_m\right)^{(1-\gamma^j)} \frac{\left(\kappa^R_{nm}\right)^{\frac{\sigma-1}{\rho}}}{\mu}\right)^{-\rho} \left[\frac{w^{j,R}_m f_m}{\sum_n \left(\frac{\kappa^R_{nm}}{P^{R,j}_n}\right)^{1-\sigma} Y_n}\right]^{\frac{\sigma-\rho-1}{\sigma-1}}}{\sum_h \lambda Y_h \left(\left(w^{j,R}_h\right)^{\gamma^j} \left(P^{j,M}_h\right)^{(1-\gamma^j)} \frac{\left(\kappa^R_{nm}\right)^{\frac{\sigma-1}{\rho}}}{\mu}\right)^{-\rho} \left[\frac{w^{j,R}_h f_h}{\sum_n \left(\frac{\kappa^R_{nh}}{P^{R,j}_n}\right)^{1-\sigma} Y_n}\right]^{\frac{\sigma-\rho-1}{\sigma-1}} + \left(\left(\omega^{j,B}_n\right)^{\gamma^j} \left(P^{j,M}_n\right)^{(1-\gamma^j)}\right)^{1-\sigma}}$$

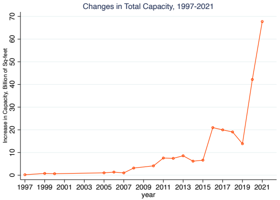
$$x^{j,B}_{nn} = \frac{\left(\left(\omega^{j,B}_n\right)^{\gamma^j} \left(P^{j,M}_n\right)^{(1-\gamma^j)}\right)^{1-\sigma}}{\sum_h \lambda Y_h \left(\left(w^{j,R}_h\right)^{\gamma^j} \left(P^{j,M}_h\right)^{(1-\gamma^j)} \frac{\left(\kappa^R_{nm}\right)^{\frac{\sigma-1}{\rho}}}{\mu}\right)^{-\rho} \left[\frac{w^{j,R}_h f_h}{\sum_n \left(\frac{\kappa^R_{nh}}{P^{R,j}_n}\right)^{1-\sigma} Y_n}\right]^{\frac{\sigma-\rho-1}{\sigma-1}} + \left(\left(\omega^{j,B}_n\right)^{\gamma^j} \left(P^{j,M}_n\right)^{(1-\gamma^j)}\right)^{1-\sigma}}$$

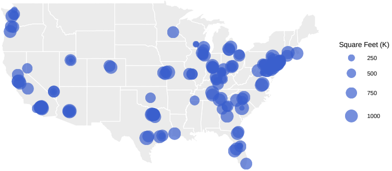
Estimation: Amazon Transportation Shock back

- Data: Amazon's Facility Network
  - address, square feet, date, type.[Houde, Newberry & Seim (HNS,2021)]
  - focus on large fulfill. & distr. centers; drop specialized, small-package









Estimation: Amazon Transportation Shock back

- Need to specify how:

$$\text{origin} \to \text{facility} \to \text{destination}$$

- HNS (2021): 90% of orders from 3 closest centers to dest.
- Assume among the 3 closest to destination, the closest to origin

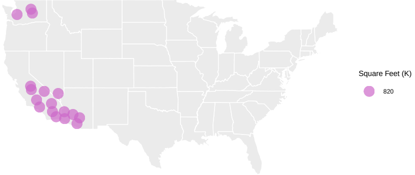
	Mean	Std. Dev.	P25	P75	Corr.
Panel A. Actual Amazon Facility					
2007	490.2	376.3	234.9	739.0	–
2017	287.9	225.6	124.7	409.0	–
Log Diff.	−0.5	0.6	−0.9	0.0	–
Panel B. Counterfactual Amazon Facility					
2007	623.4	400.3	349.6	897.4	0.10
2017	335.2	278.4	143.9	412.1	0.58
Log Diff.	−0.7	0.8	−1.1	0.0	−0.02

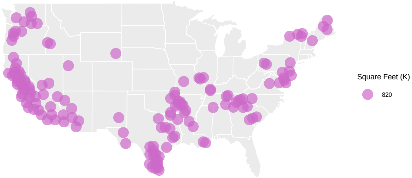
Estimation: Amazon Transportation Shock back

- Spatial Simulated IV
  - concern: endogeneity of facilities
  - simulate facilities' locations based only on geo. cost factors, to be uses as IV (Duflo et.al, 2007; Lipscomb et.al, 2013; Faber 2014)
  - need orthogonality of geo. factors
- Simulation Steps
  - based on observed # of new centers, determine AMZ's budget
  - rank counties by geo. factors
  - highest ranks get new centers

		Dependent: $1\{$AMZ Center$\}$
Temperature (Lag)	Mean	-0.011
	Minimum	-0.002
	Maximum	0.046***
Precipitation (Lag)	Mean	-0.032
	Minimum	0.043
	Maximum	-0.015
Elevation	Mean	-0.001***
	Minimum	0.000
	Maximum	0.001***
Tornado	Magnitude	-0.051
	Injuries	-0.110
Year FE		X
Observations		55,259
Psudo R-squared		0.1663

Estimation: Amazon Transportation Shock back



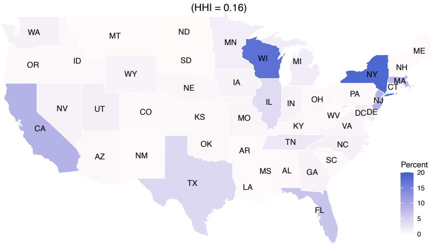


	Dependent (distance in Log)	
	Actual	Counterfactual
First Stage Results		
Counterfactual	0.40***	
	[0.02]	
F-Stats	670	
Robustness		
Avg. lag GDP		0.00
		[0.00]
Avg. GDP growth		-0.00***
		[0.00]
Observations	4,704	2,352
R-squared	0.12	0.04

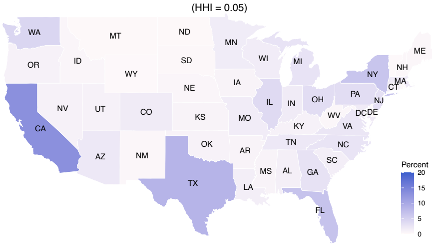
Empirical Patterns back

- 1a: Online retail sales are more concentrated than average retail sales...

(1) Regional Share of E-commerce Sales



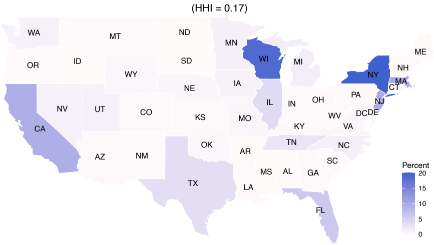
(2) Regional Share of Retail Sector Sales



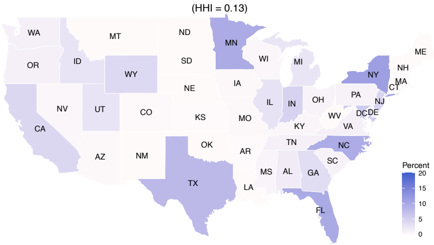
Empirical Patterns back

- 1a: Online retail sales are more concentrated than average retail sales...
- 1b: …and those that are FBA more concentrated than non-FBA

(3) Regional Share of E-commerce Sales with FBA



(4) Regional Share of E-commerce Sales without FBA



Appendix

Empirical Patterns back

- 2: Durable/standardized ones are less concentrated

Table: HHI Index by Product Categories

Category name	HHI Index
Toys & Games	0.12
Patio, Lawn & Garden	0.12
Arts, Crafts & Sewing	0.07
Sports & Outdoors	0.14
Office Products	0.16
Grocery & Gourmet Food	0.08
Tools & Home Improvement	0.21
Movies & TV	0.08
Musical Instruments	0.10

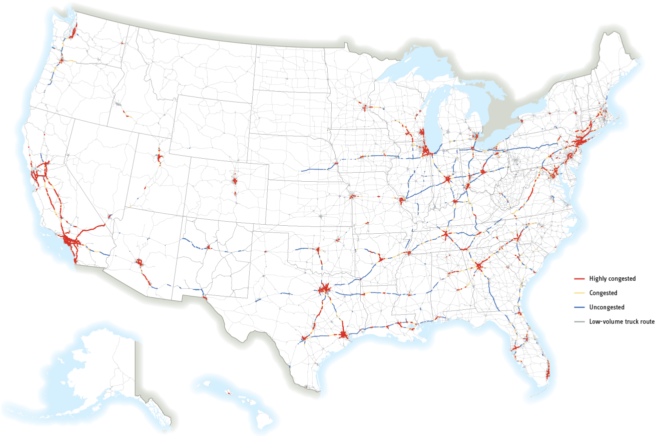
Empirical Patterns back

- 3a: Online retail is less correlated with population or taxes

Dependent Variable (in %)	Online Retail	Overall Retail
$\ln$ (corporate tax)	-0.01	0.03*
	[1.29]	[0.02]
Population share (%)	14.54*	1.06***
	[7.92]	[0.26]
Year, State FE	X	X
Observations	230	230
R-squared	**0.52**	**1.00**

Empirical Patterns back

- 3a: Online retail is less correlated with population or taxes
- 3b: ...and the concentration aligns with truck routes



Appendix

Summary back

- Environment
  - N regions: n (destination), m (origin)
  - J sectors: j (home production, service) & (durable, non-durable)
  - 3 subsectors: M (manufacturer), R (online retailer), B (brick-and-mortar)

1. **Demand**: Sequential directed search → CES w/. demand shifter
2. **Intermediate**: Ricardian (EK) → manuf. trade flow
3. **Online Seller**: Location choice → agglomeration, retail trade flow
   - Two approaches: Arkolakis et al. (2018, 2017) vs. Chaney (2008)
   - Key difference: multiple destinations & origins, vertical production
4. **Worker**: Roy labor supply

Consumer Demand

- Sequential Directed Search
  - A continuum of consumers (n), sector share (η^j)
  - Pick 1 among measure $1 + O^j$ sellers, $O^j = \sum_m O^j_m$
  - $v^j_{nm} = \ln \eta^j y_n - \ln p^{j,K}_{nm} + \epsilon^{j,K}_{nm}$ (i.i.d. $E(\epsilon^{j,B}_{nm}) = 0$, and $E(\epsilon^{j,R}_{nm}) = ln(\mu)$)
  - Sequential directed search: pay k to see $\epsilon^{j,K}_{nm}$, or continue Weitzman ('79)

1. Any SDM has a discrete choice model (DCM) w/. same demand proof
2. CES demand is a special case of DCM with extreme type I error proof

Theorem

A rep. consumer in n with weights η^j has nest CD-CES demand as below under sequential ordered search and if $\epsilon^{j,K}_{nm}$ is distributed extreme type I

$$C_n = \Pi^J_{j=1}(C^j_n)^{\eta^j}, \quad C^j_n = [(c^B_{nn})^{\frac{\sigma^j-1}{\sigma^j}} + \mu \sum_{m=1}^{N} \int_0^{O^j_m} (c^R_{nm}(i))^{\frac{\sigma^j-1}{\sigma^j}} di]^{\frac{\sigma^j}{\sigma^j-1}}$$

Vertical Production

- Intermediate Varieties (M)
  - A rep. firm in (n, j, M) produces varieties $e^j \in [0, 1]$

$$q_n^{j,M}(e^j) = a_n(e^j) l_n(e^j)$$

- Retail Sector (R/B)
  - Collect varieties $e^j \in [0, 1]$: $q_n^{j,R/B} = [\int_0^1 q_n^{j,M}(e^j)^{\frac{\alpha^j - 1}{\alpha^j}} d\phi^j(a^n(e^j))]^{\frac{\alpha^j}{\alpha^j - 1}}$

$$Q_n^{j,R/B} = z_n^{j,R/B} \left[(h_n^{j,R/B})^{\beta_n} (l_n^{j,R/B})^{1-\beta_n} \right]^{\gamma_n^j} \left[q_n^{j,R/B} \right]^{1-\gamma_n^j}$$

  - i.i.d. Fréchet (θ^j, T_n^j). Intermediate exp. share: $x_{nm}^{j,M} = \frac{(\kappa_{nm}^M c_m^{j,M})^{-\theta^j} T_m^j}{\sum_{g=1}^N (\kappa_{ng}^M c_g^{j,M})^{-\theta^j} T_g^j}$
  - Unit cost: $c_n^{j,R/B} = (\omega_n^{j,R/B})^{\gamma_n^j} (p_n^{j,M})^{1-\gamma_n^j} / z_n^j$. For online: $p_{nm}^{j,R} = c_m^{j,R} \kappa_{nm}^R$

Online Retailer Location

- Optimal Location (R) alternative
  - Online retailers draw $(z_1^{j,R}, ..., z_N^{j,R})$, entry cost f_m. Optimal location:

$$m^* = \arg\min_m \left\{ \sum_n \left(\tilde{\sigma} \frac{c_m^{j,R}}{z_m^{j,R}} \frac{\kappa_{nm}^R}{P_n^{j,R}} \right)^{\sigma^j - 1} \cdot \frac{1}{\eta^j X_n} \right\} \quad \left(\equiv \frac{\tilde{\sigma}\xi_m^j}{z_m^{j,R}} \right)$$

Entry: $\sum_n \left(\frac{p_{nm}^{j,R}/\mu}{P_n^{j,R}} \right)^{1-\sigma^j} \eta^j X_n \geq \sigma^j w_m^{j,R} f_m$. Thold: $\bar{c}_m^{j,R} = \frac{\mu z_m^{j,R}}{\tilde{\sigma}^j} \left[\frac{\sigma^j}{\eta^j} \frac{w_m^{j,R} f_m}{\sum_n (\kappa_{nm}^R / P_n^{j,R})^{\sigma^j - 1} X_n^{-1}} \right]^{\frac{1}{1-\sigma^j}}$

- Aggregate Retail Trade
  - Multi-var Pareto : $P(Z_1^j < z_1, ..., Z_N^j < z_N) = 1 - \left(\sum_{m=1}^N [T_m^{j,R} z_m^{-\phi}]^{\frac{1}{1-\rho}} \right)^{1-\rho}$

$$\Psi_m^j = P\left(m = argmin_m \left\{ \frac{\tilde{\sigma}\xi_m^j}{z_m^j} \right\} \cap c_m^{j,R} < \bar{c}_m^j \right) = \psi_m^j (\bar{c}_m^j)^\phi \quad \psi_m^j = \frac{T_m^{j,R} (\xi_m^j)^{\frac{-\phi}{1-\rho}}}{\sum_{m=1}^N [T_m^{j,R} (\xi_m^j)^{-\phi}]^{\frac{-\rho}{1-\rho}}}$$

Bilateral online retail exp. share	Regional brick-and-mortar exp. share
$x_{nm}^{j,R} = \frac{\Psi_m^j (\kappa_{nm}^R c_m^{j,R} / \mu)^{1-\sigma}}{\sum_h \Psi_h^j (\kappa_{nh}^R c_h^{j,R} / \mu)^{1-\sigma} + \frac{1}{O} (c_n^{j,B})^{1-\sigma}}$	$x_n^{j,B} = \frac{\frac{1}{O} (c_n^{j,B})^{1-\sigma}}{\sum_h \Psi_h^j (\kappa_{nh}^R c_h^{j,R} / \mu)^{1-\sigma} + \frac{1}{O} (c_n^{j,B})^{1-\sigma}}$

Labor Supply

- Employment Share
  - L_n HHs choose sector $\{j, K\}$ (home production $j = 0$)
    - $K = \{M, R, B\}$ the three subsectors for dur/non-dur sectors, $\varnothing$ for others
  - Draw $z_n^{j,K}$ from i.i.d. Fréchet $(\nu_n, A_n^{j,K})$

$$\pi_n^{j,K} = \frac{A_n^{j,K}(w_n^{j,K})^{\nu_n}}{\Phi_n}, \text{ where } \Phi_n = \sum_{j=0}^{J} \sum_{K=\{M,R,B,\varnothing\}} A_n^{j,K}(w_n^{j,K})^{\nu_n}$$

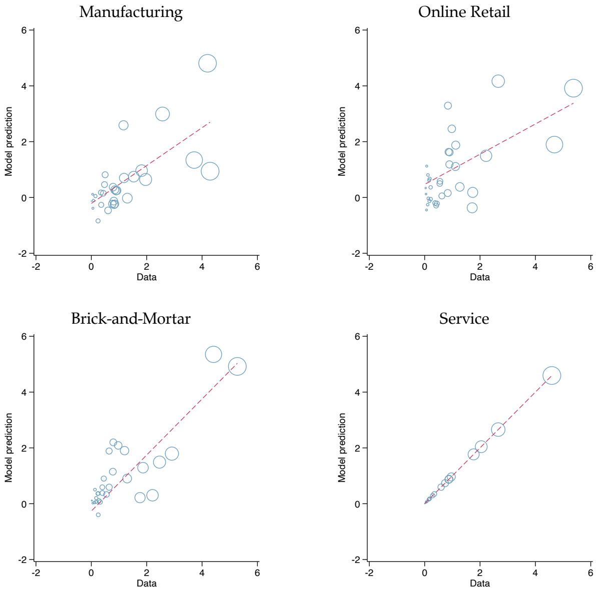
- Sectoral Wage Income
  - Let $l_n^{j,K}$ efficiency units of labor provided to sector (j, K)
  - Wage income in (j, K) becomes $w_n^{j,K} l_n^{j,K} = \Gamma(\frac{\nu_n - 1}{\nu_n})\Phi_n^{1/\nu_n} \pi_n^{j,K} L_n$

Calibration: General back

Section	Param.	Description	Estimation/Caliberation
Consumer	η_n^j	Sector share of consumption	CFS 2007
	σ^j	Elasticity of subs. across retailers	Keepa + IV
Labor Supply	π_n^j	Share of empployment	CBP, ACS
	υ^n	Fréchet shape of worker product.	Galle, Rodríguez-Clare & Yi (2022)
Production	β_n^j	Share of structures	BEA + Greenwood et. al (1997)
	θ^j	Fréchet shape of sector product.	Caliendo and Parro (2015)
	γ_n^j	Value-added share of retail goods	BEA, CFS
Expenditure	$x_{ni}^{j,M}$	Interm. expenditure share	CFS 2007
	$x_n^{j,B}$	Brick-and-motar expenditure share	CFS 2007, E-Stats
	$x_{nm}^{j,R}$	E-commerce expenditure share	CFS 2007, E-Stats
	$p_n^{j,B}$	Brick-and-motar price index	CFS 2007, E-Stats, CES

Baseline Economy (2007): Model vs. Data

- Model implied regional income (untargeted)



Sequential Estimation: Amazon Shock back

Section	Param.	Description	Estimation/Caliberation
	$\hat{\kappa}^R_{nm}$	Iceberg cost change	Amazon data + CFS 2007 + IV
Amazon	μ	Matching efficiency	E-stats + CES
Shock	Ψ^j_m	Online retailer location probability	Keepa
	O	Measure of online retailers	E-stats
	T^j_n	Fréchet scale of sectoral product.	Assume constant
	A^j_n	Fréchet scale of labor product.	Assume constant

Sequential Estimation: Amazon Shock

- Extrapolate Amazon Ice-berg cost shock
  - **Intuition**: Ice-berg is increasing in distance
  - Estimate coefficient of ice-berg cost on shipping distance details

$$ln(\kappa^{j,R}_{nm}) = \delta^j \text{Distance}_{nm} + X'_{nm}\theta + \delta^j_n + \delta^j_m + \epsilon^j_{nm}$$

  - Estimate reduction in shipping distance due to Amazon
    - Build counterfactual facilities based on exog. factors as IV for actual ones
- Back-out online matching efficiency
  - **Intuition**: % online exp. should inform matching, conditional on shipping

$$\sum_{m=1}^{N} x^{j,R}_{nm} / x^{j,R}_{nn} = (\mu)^{\sigma^j - 1} \sum_{m=1}^{N} M_m (p^{j,R}_m \kappa^R_{nm} / p^{j,R}_{nn})$$

    - Use Keepa for M_m, above estimated κ^R_{nm}, CES for $p^{j,R}_m$, $p^{j,R}_{n0}$

δ^{dur}	δ^{nondur}	$\hat{\kappa}$	μ
1.5	2.1	0.97	1.27
[0.2]	[0.6]	[0.15]	[1.46]