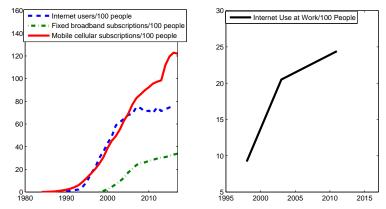
Geographic Fragmentation in a Knowledge Economy: Theory and Evidence from the US

Yang Jiao (FISF, Fudan University) Lin Tian (INSEAD & CEPR)

May 2021

Motivation

- Revolutionary development in ICT, e.g., Internet technology
- In the United States



Data Source: the WDI data in the left panel; authors' calculation from Current Population Survey Internet and Computer Use Supplement in the right panel

Motivation

- Profound impacts on the labor market: geographic fragmentation, e.g., sourcing, headquarter-subsidiary relation
- Increasingly fragmented production process across geographic boundaries:
 - internationally: a huge literature (e.g., Hummels, Ishii and Yi, 2001; Antras, Garicano & Rossi-Hansberg, 2006; Grossman & Rossi-Hansberg, 2008; ...)
 - domestically: underexplored
 - * focus of this paper
- Domestic fragmentation:
 - Quantitatively important: e.g., 95% of sourcing done domestically (BCG survey, 2015)
 - Labor mobile across regions
 - \star Spatial movement of economic activities \implies Inter-regional redistribution of skills
- Research question:

How the rise in *cross-region productions*, driven by internet improvement, shapes the distribution of skills across US cities and welfare?

Preview of Results

- Stylized facts:

 - Industries tend to fragment more see larger increases in spatial skill segregation
- A spatial eqm model of production fragmentation + heterogeneous agent
 - knowledge (skilled) + standardized production (unskilled)
 - ↓ communications costs: ↑ cross-city joint production, ↑ skilled share in larger cities; and ↓ skilled share in smaller cities
- Empirical support for model predictions on Internet improvement and skill flows
- Quantitatively evaluate the importance of proposed mechanism

Related Literature

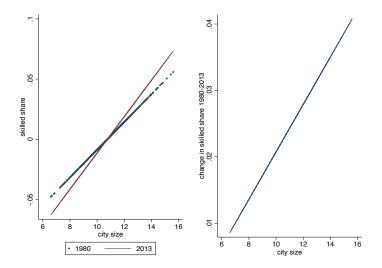
- International offshoring: Feenstra (1998); Hummels, Ishiii & Yi (2001); Antras, Garicano & Rossi-Hansberg (2006); Grossman & Rossi-Hansberg (2008)
- Domestic production fragmentation: Duranton & Puga (2005); Liao (2012); Santamaria (2018); Eckert (2019); Acosta (2020); Hsieh & Rossi-Hansberg (2020)
- Spatial equilibrium model in a system-of-cities: Behrens, Duranton & Robert-Nicoud (2014); Davis & Dingel (2012, 2019)
- Quantitative spatial equilibrium analysis: Allen & Arkolakis (2014); Allen, Arkolakis & Takahashi (2014)
- Impact of ICT technology on production organizations: Fort (2017); Tian (2019)

Stylized Facts

Data and Definitions

- Census Integrated Public Use Micro Samples (IPUMS):
 - ▶ 1980: 5 percent census; 2011-2013 three-year ACS: 3 percent sample
 - Individuals between age 16 and 64
- Local labor markets: 722 commuting zones
- City sizes: total labor supply (robust if population)
- Two skill groups (occupation based): high (mean wage rank above 75%); low (others)
 - robust to other thresholds: 80% or 67%
 - robust to education: COL+

Skilled Empl Share & City Size



Notes: the left panel displays the regression line for the high skilled share (demeaned) in 1980 and 2013 against city size (log of 1980 population). The right panel displays the change in the skilled share from 1980 to 2013.

Change in Skilled Empl Share and City Size: 1980-2013

Dependent variable: change in the skilled share				
	(1)	(2)		
City Size	0.004***	0.005***		
	(0.001)	(0.001)		
State fixed effect	No	Yes		
Observations	722	722		
R^2	0.037	0.357		

* p < 0.10, ** p < 0.05, * * * p < 0.01

Robust standard errors are in parentheses. Significance levels: * 10%, ** 5%, ***1%.

• Larger cities become increasingly specialized in skill-intensive activities

Spatial Skill Segregation: 1980 - 2013

• Kremer & Maskin (1996) segregation index

$$\rho = \frac{1}{5} \sum_{s} \left[\frac{\sum_{c} N_{cs} \cdot (\pi_{cs} - \pi_{s})^{2}}{N_{s} \cdot \pi_{s} \cdot (1 - \pi_{s})} \right]$$

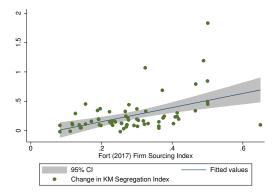
where

- N_{cs}: employment in sector s and city c
- ► N_s: total sectoral employment
- $\pi_{cs} = \frac{N_{cs}^{skilled}}{N_{cs}}$: high skilled employment share in sector *s* and city *c* • $\pi_s = \frac{N_s^{skilled}}{N_c}$: high skilled employment share in sector *s*
- Larger ρ : greater extent of segregation
- KM index more than tripled from 1980 2013

Year	ρ	95% Confidence Interval
1980	0.00746	(0.00741, 0.00752)
2013	0.0204	(0.0202, 0.0205)

Production Fragmentation and Spatial Segregation

• Change in sector-level KM index and Fort (2017) sourcing index



Notes: each dot represents an NAICS4 industry. The correlation between change in KM skill segregation index and Fort sourcing index is 0.47.

Industries that source more are also those that tend to undergo greater skill segregation

Summary of Stylized Facts

- Three stylized facts:
 - 1. Larger cities have a comparative advantage in skill-intensive activities (Glaeser, 2008; Davis & Dingel, 2014)
 - 2. Pattern of specialization has become stronger, as skilled and unskilled workers become more segregated geographically
 - 3. Extent of segregation strongly associated with production fragmentation
- Central hypothesis
 - ▶ ICT (e.g., Internet technology) improvement reduces communication frictions:
 - \implies \uparrow geographic fragmentation (cross-city joint productions)
 - \implies Reinforce initial patterns of specialization
 - \implies Spatial redistribution of skills

Theory

Set-up and Preferences

- Finite number of cities $n \in \mathcal{N}$, with exogenous housing supply H_n
- Continuum of agents, distinguished by their skill levels
 - L^m skilled labor (called "managers")
 - L^p unskilled labor (called "production workers")
- Agents inelastically supply labor, mobile across n
- Utility function:

$$U(x,h) = \alpha^{-\alpha} (1-\alpha)^{-(1-\alpha)} x^{\alpha} h^{1-\alpha},$$

where

- x: homogeneous good
- h: housing

Production

• Managers who live in city *n* may hire workers in any city *c* with:

$$y_{nc} = a_{nc} \cdot I_c^{\beta}, \ \beta < 1$$

• *a_{nc}*: manager's productivity

$$a_{nc} = f(L_n^m) imes ar{a}_{nc}$$

- $f(L_n^m) = (L_n^m)^{\gamma}$, $\gamma \ge 0$: agglomeration externalities in n.
- \bar{a}_{nc} : a random draw
 - ★ Manager living in city *n* draws \bar{a}_{nc} for all $c \in \{1, ..., N\}$ cities
- ā_{nc} follows Fréchet distribution

$$G(a) = \exp\left(-T_n a^{-\theta}\right)$$

- *T_n*: exogenous technology parameter in *n*
- $\theta > 0$: dispersion of manager's productivities across cities

Manager's problem

• Income of a manager living in *n*, hiring workers in *c*:

$$\pi_{nc} = \frac{a_{nc}}{\tau_{nc}} I^{\beta} - w_c I = \beta^{\frac{\beta}{1-\beta}} (1-\beta) (\frac{a_{nc}}{\tau_{nc} w_c^{\beta}})^{\frac{1}{1-\beta}}$$

- τ_{nc} ≥ 1: "fragmentation" costs, e.g., cross-city communication, off-site coordination, search frictions...
- Given Fréchet productivity assumption, fragmentation gravity equation:

$$x_{nc} \equiv rac{L_{nc}^m}{L_n^m} = rac{T_n (au_n w_c^eta)^{- heta}}{\Phi_n},$$

where $\Phi_n \equiv \sum_k T_n (\tau_{nk} w_k^\beta)^{-\theta}$ ("fragmentation potential" of city *n*)

- *x_{nc}*: share of managers in *n* producing in *c*
- Expected income of a manager living in n:

$$E[\pi_n] = \zeta[[f(L_n^m)]^{\theta} \Phi_n]^{\frac{1}{\theta(1-\beta)}}$$

Spatial Indifference Conditions

- In equilibrium, agents indifferent to living locations
- Production workers' indifference condition:

$$\frac{w_n}{p_n^{1-\alpha}} = \frac{w_{n'}}{p_{n'}^{1-\alpha}}, \quad \forall n, n'$$

• Managers' Indifference condition:

$$\frac{E[\pi_n]}{p_n^{1-\alpha}} = \frac{E[\pi_{n'}]}{p_n'^{1-\alpha}}, \quad \forall n, n'$$

Equilibrium Properties

- WLOG, suppose $T_n > T_c$
- Internet Autarky: $\tau_{nc} = \infty$
 - Cities with higher T_n are larger: more L_n^m and L_n^p
 - Skilled share the same across cities
- Internet Openness: $\tau_{\it nc} < \infty$
 - $\tau_{nc} = \tau_{cn} \downarrow$ locally.
 - L_n^m and $L_c^p \uparrow$; L_c^m and $L_n^p \downarrow$
 - Skilled shares: \uparrow in n; \downarrow in c
 - Stronger agglomeration externality, $(L_n^m)^\gamma \implies$ Greater labor relocation for both managers and low-skilled workers

Infinite Fragmentation Cost

- Set $f(L_n^m) = (L_n^m)^{\gamma}$ and assume $\gamma + 1 > \frac{\gamma}{1-\alpha}$, when $\tau_{nc} \to +\infty$, $\forall n \neq c$, the spatial equilibrium exists and is unique
- The number of managers in each city L_n^m and the number of production workers in each city L_n^p

$$L_n^m \propto T_n^\kappa$$

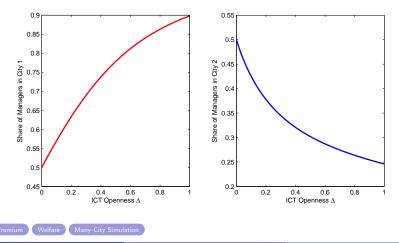
 $L_n^p \propto T_n^\kappa$,

where
$$\kappa = rac{1}{1-lpha} - 1 \over 1+\gamma - rac{\gamma}{1-lpha}} rac{1}{ heta} > 0$$

• Skilled share $L_n^m/(L_n^m + L_n^p)$ in each city the same across all cities

Two-City Simulation: Skilled Shares

- Internet Openness $riangle_{nc} = au_{nc}^{- heta}$
 - $\triangle_{nc} \uparrow$
 - ▶ Suppose $T_1 > T_2$, then $L_1^m/(L_1^m + L_1^p) \uparrow$ and $L_2^m/(L_2^m + L_2^p) \downarrow$



Heterogeneous effects of Internet

• Heterogeneous effects of Internet on city skill composition

- ► Large city: ↑ share of skilled workers
- Small city: \downarrow share of skilled workers
- Empirical specification:

 $\Delta \text{skilled share}_i = \beta_0 + \beta_1 \text{ city size}_i + \beta_2 \Delta \text{internet}_i + \beta_3 \text{ city size}_i * \Delta \text{internet}_i$ $+ \gamma X_i + \epsilon_i$



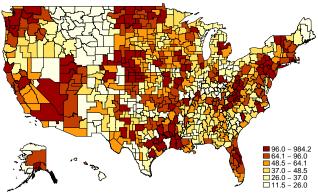
- State FEs
- Telephone penetration rate in 1980

• $\beta_2 < 0$, $\beta_3 > 0$

Empirical Support

US Internet Infrastructure

- Data source: US Federal Communications Commission
- Block-level Internet download and upload bandwidths 2014
 - fixed broadband suppliers file Form 477 on maximum bandwidths
 - Population-weighted average CZ-level measures



Notes: Speeds are measured in Megabytes per second.

Identification Challenges

 $\begin{aligned} \Delta \text{skilled share}_i = & \beta_0 + \beta_1 \operatorname{city size}_i + \beta_2 \Delta \text{internet}_i + \beta_3 \operatorname{city size}_i * \Delta \text{internet}_i \\ & + \gamma X_i + \epsilon_i \end{aligned}$

- 1. Long-run local employment trends
- 2. Unobserved local shocks affecting both internet quality and changes in skill share

3. Reverse causality: local labor demand shocks drive internet provision

Identification Challenges

 $\begin{aligned} \Delta \text{skilled share}_i = & \beta_0 + \beta_1 \operatorname{city size}_i + \beta_2 \Delta \operatorname{internet}_i + \beta_3 \operatorname{city size}_i * \Delta \operatorname{internet}_i \\ & + \gamma \mathsf{X}_i + \epsilon_i \end{aligned}$

- 1. Long-run local employment trends
- 2. Unobserved local shocks affecting both internet quality and changes in skill share
 - (Large Cities) Omitted variable: + skilled share and + internet
 - (Small Cities) Omitted variable: skilled share and + internet
- 3. Reverse causality: local labor demand shocks drive internet provision
 - ► (Large Cities) Larger skilled share ⇒ internet improvement
 - ► (Small Cities) Smaller skilled share ⇒ internet improvement

Identification Strategies

- Falsification test to rule out long-run trends: replacing LHS by change in employment share between 1950 1980
- Instrumentation strategy to address OVB and reverse causality,
 - Instrument: Average elevation of the local terrain (Jaber, 2013; Amorim, Lima & Sampaio, 2015)
 - Higher elevation areas less costly for broadband infrastructure deployment and maintenance
 - e.g., proneness to flooding, summer temperature



Results

Dependent variable: change in the share of high-skill employment						
	1980-2013			1950-1980		
	0	LS	2SLS	OLS		2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Internet quality	023**	029**	138***	005	012	013
	(.009)	(.012)	(.034)	(.017)	(.020)	(.037)
Internet quality $ imes$ city size	.0022**	.0028**	.0119***	-0.000	.001	.001
	(.0008)	(.0011)	(.003)	(.001)	(.001)	(.003)
State Fixed Effect	No	Yes	Yes	No	Yes	Yes
Observations	722	722	722	722	722	722
R^2	.045	.360	.253	.048	.284	-0.338
S-W F-stats (First Stage)						
Internet quality			12.92			12.92
Internet quality \times city size			11.15			11.15

Notes: City size is measured by log(population in 1980) and is always included as a control variable. Standard errors are in parentheses. Robust standard errors are used when there is no state fixed effect. Standard errors clustered at the state level when there is state fixed effect. We also report Sanderson-Windmeijer (S-W) F- statistics for the first stage regressions. * p < 0.10, ** p < 0.05, *** p < 0.01

Dealing with "Exclusion Restriction"

- Concern:
 - \blacktriangleright Theory: Internet \implies Production fragmentation \implies Skill relocation
 - ▶ Empirics: Internet ⇒ Skill relocation
- Some industries are more likely to fragment (Fort, 2017)
 - Cities with greater concentration of such industries would undergo greater extent of skill relocation
- Separate cities into two groups based on average fragmentation intensities

$$\sum_{i} \frac{\text{Sourcing Index}_i \times L_{ic}}{L_c}$$

- Repeat the IV regression separately for the two groups of cities
- Hypothesis: fragmentation-intensive cities would experience more skill relocation

Results

Dependent variable: change in share of high-skill employment				
	Baseline	Non-Fragment-Intensive	Fragment-Intensive	
	(1)	(2)	(3)	
Internet quality	138***	102	139*	
	(.034)	(.062)	(.077)	
Internet quality $ imes$ city size	.0119***	.0084	.0122**	
	(.0031)	(.0062)	(.0057)	
State Fixed Effect	Yes	Yes	Yes	
Observations	722	361	361	
R ²	.253	.233	.342	
S-W F-stats (First Stage)				
Internet quality	12.92	11.02	18.09	
Internet quality \times city size	11.15	10.01	24.53	

Notes: City size is measured by log(population in 1980) and is always included as a control variable. Standard errors are in parentheses. Robust standard errors are used when there is no state fixed effect. Standard errors clustered at the state level when there is state fixed effect. We also report Sanderson-Windmeijer (S-W) F- statistics for the first stage regressions. * p < 0.10, ** p < 0.05, *** p < 0.01

Quantitative Analysis

Model Extension

• Housing market (Ganong and Shoag, 2017; Giannone, 2019)

$$H_n = \bar{H}_n p_n^{\mu} \implies p_n = \left(\frac{(1-\alpha)W_n}{\bar{H}_n}\right)^{\frac{1}{\mu+1}}$$

- More extensions (in progress):
 - Endogenous amenity (Diamond, 2016)
 - * Differences in valuations of amenities between skill types

$$U_p = c^{\alpha} h^{1-\alpha}$$
$$U_m = c^{\alpha} h^{1-\alpha} A^{\zeta}$$

* Amenity supply

$$\log A_n = \kappa (\log L_n^m - \log L_n^p)$$

Skill-biased technical change

Assigning Parameter Values

- Literature
 - Share of spending on housing: (1 α) = 0.24 (Davis & Ortalo-Magne, 2011; Behrens, Duranton & Robert-Nicoud, 2014)
 - Span of control: $\beta = 0.53$ (Buera & Shin, 2013)
 - Housing supply elasticity: $\mu = 0.135$ (Giannone, 2019)
- Strength of agglomeration γ : match elasticity of wage w.r.t. city size
- Dispersion of manager's productivity θ : match high skilled hourly wage distribution
- Housing supply \overline{H}_n : match city-level wage and total income
- City technology T_n : match city-level differences on wage, size, and fragmentation potential Φ_n

Skip to Counterfactual

Bilateral Fragmentation Costs

• Semi-parametric form:

$$\log \tau_{\mathit{nc}} = \delta^{\mathit{d}} \log \mathit{d}_{\mathit{nc}} + \delta^{\mathit{l}} \mathit{q}_{\mathit{nc}} + \lambda_{\mathit{nc}}$$

- Power functions of bilateral distance d_{nc} , and internet connectivity $q_{nc} = q_n \times q_c$
- Other bilateral costs λ_{nc}
- Infer τ_{nc} from fragmentation gravity equation
 - Recall:

$$X_{nc} = L_n^m \frac{T_n \tau_{nc}^{-\theta} w_c^{-\beta\theta}}{\Phi_n}$$

Taking ratios, we get:

$$\tau_{nc} = \left(\frac{w_c^{\beta\theta} X_{nc}}{w_n^{\beta\theta} X_{nn}}\right)^{-1/\theta}$$

- X_{nc} : number of subsidiaries in c belonging to headquarters in n
 - * Obtained from Orbis Database More details Orbis database

Elasticity Estimates

Estimate

$$\log \tau_{nc} = \delta^d \log d_{nc} + \delta^l q_{nc} + \lambda_{nc}$$

using:

$$\log \tau_{nc} = \chi_n + \iota_c + \delta^d \log d_{nc} + \delta^I q_n q_c + \Theta \mathsf{H}_{nc} + \varepsilon_{nc}$$

where H_{nc} :

- Same state
- Shared border
- Racial affinity

Estimates	OLS	PPML
$\hat{\delta^d}$.134***	.231***
	(.004)	(0.006)
$\hat{\delta}'$	010***	010***
	(.0027)	(.006)
City Fixed Effects	Yes	Yes
Controls	Yes	Yes
Ν	44,203	505,008

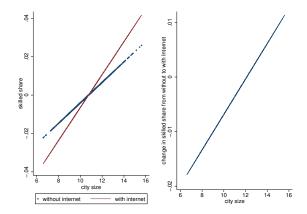
Notes: Robust standard errors in parentheses. Significance levels: * 10%, ** 5%, ***1%.

Role of the Internet in Skill Redistribution

• Assume no internet quality improvement between 1980 and 2013

$$\log \tilde{\tau}_{nc} = \log \tau_{nc} - \hat{\delta}^{I} q_{nc}$$

• Solve for counterfactual skilled share in each city.



Skill Redistribution and Welfare

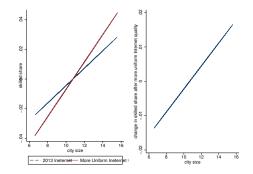
- Slope coefficient: \triangle 0.0031**
 - > Without internet, the observed skill redistribution in the US would have been reduced by about 0.00310/0.00503 = 61%
- Welfare effects
 - Unskilled: 3.88%
 - Skilled: 3.66%

	Δ Managers' Welfare	Δ Workers' Welfare
Direct Effect	2.93%	2.89%
GE Effect	0.95%	0.77%
Total	3.88%	3.66%

Table: Decomposition of Welfare Changes

Narrowing the Digital Divide

- Policy experiment: programs improving internet quality in less connected places, e.g., Connect America Fund
 - Upgrade internet in cities with below-median level to the median level



- ► More spatial skill relocation: semi-elasticity of skill share wrt city size ↑ .0007
- ▶ Welfare implications: Unskilled +0.19%; Skilled +0.17%

Conclusion

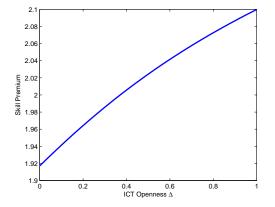
• Document facts using US individual level data

- Larger cities disproportionately attract the skilled 1980-2013
- More skill segregation occurs in fragmentation intensive industries
- Develop a model of domestic production fragmentation with heterogeneous skills
 - skill distribution
 - communication cost, city size and skill flows
- Provide empirical support for key model predictions
- Quantify the importance of domestic production fragmentation, aided by Internet improvement, in shaping the spatial skill distributions and welfare

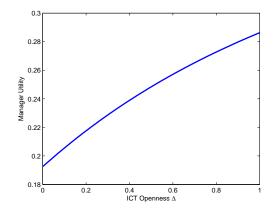
Appendix

Two-City Simulation: Skill Premium

Skill Premium = log
$$E[\pi_n] - \log w_n = \frac{1}{1-\beta} \log f(L_{mn}) + \frac{1}{(1-\beta)\theta} \log \Phi_n - \log w_n$$

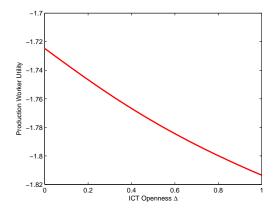


N=2 Simulation: Welfare for Managers



Back

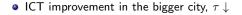
N=2 Simulation: Welfare for Production Workers

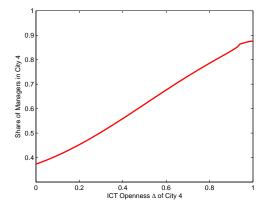


Back

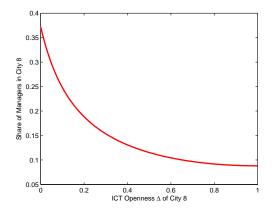
Many-City Simulation: N=8

• Four big and four small: $T_1 = T_2 = T_3 = T_4 > T_5 = T_6 = T_7 = T_8$





ICT Improvement In One Small City



Back

Orbis Database

- Orbis: shareholders with strictly more than 50% ownership
- map zip code to CZ using Missouri Census Geocorr
- count the number of bilateral headquarter-subsidiary



Joint-production from Orbis Database

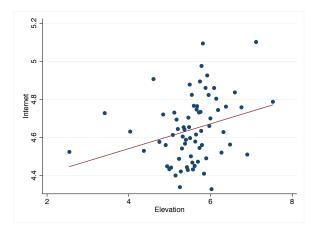
- X_{nc} : number of subsidiaries in c belonging to headquarters in n
- Limitations: not all cross-city productions are captured
- Reasonable starting point:
 - Fits skilled-unskilled production setting in the theory well
 - Identifies one specific channel through which firms can achieve fragmented production



Instrument Variable

- Cable infrastructure prone to damage from flooding, high ground temperatures, and excessive precipitation (Zimmerman and Faris, 2010)
- Land elevation are strongly correlated with these natural conditions
 - ▶ Greater flooding risk (Michel-Kerjan et al., 2010, Landry and Parvar, 2011)
 - \implies the need to safeguard broadband facilities from being submerged under water (Norhaus, 2010; Rosenzweig et al., 2011)
 - \implies difficulty of burying cable underground (Bascom and Antoniello, 2011)
 - Higher summer temperature (Willmott and Matsuura, 1995)
 - \implies higher installation costs, e.g., additional cables, artificial soil to absorb heat (Daly et al., 2008)
- Heavily associated with the use of cable technology: 90% of the broadband market in the US
 - ADSL technology in other countries, e.g., Western Europe (Jaber, 2013)

Zero Stage Results



Internet_n =
$$\alpha_1 + \alpha_2 Elevation_n + X_n + \epsilon_c$$

 $\hat{\alpha}_2 = .065(.029)$

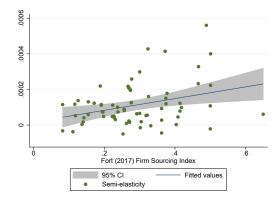
Does Internet Improve Trade in Goods

Dependent variable	log(shipment)	log(shipment)
	(1)	(2)
log (distance)	-1.236***	-1.239***
	(.0026)	(.027)
$q_i * q_j$.058	.039
	(.094)	(.053)
\boldsymbol{q}_i	.489	
	(.349)	
q_j	.379	
	(.356)	
Fixed Effects	No	Yes
N	4801	4801

Table: Gravity Equation Estimates for Trade in Good

Notes: Robust standard errors in parentheses. Significance levels: * 10%, ** 5%, ***1%.

Production Fragmentation and Spatial Segregation



Notes: each dot represents an NAICS4 industry.