Geographic Fragmentation in a Knowledge Economy: Theory and Evidence from the United States*

Yang Jiao† and Lin Tian‡

September 9, 2024

Abstract

This paper explores how internet technology advancements drive cross-city collaborations (or "geographic fragmentation"). We use a spatial equilibrium model with cross-city production and skill heterogeneity to analyze the effects of reduced communication costs on domestic fragmentation. Our model suggests that better internet leads to increased production fragmentation, concentrating skilled workers in larger cities and reducing their numbers in smaller ones. Empirical validation using a novel instrumental variable approach confirms these predictions. Our calibrated model indicates that internet advancements have increased real wages for both high- and low-skill workers, with welfare improvements partly due to spatial reorganization from enhanced production fragmentation.

*We are indebted to Treb Allen, Costas Arkolakis, Davin Chor, Donald Davis, Jonathan Dingel, Jonathan Eaton, Fabian Eckert, Pablo Fajgelbaum, Teresa Fort, Juan Carlos Hallak, Jefferey Lin, Dan Lu, Rocco Macchiavello, Andrew H. McCallum, Antonio Miscio, David Nagy, Stephen Redding, Stephanie Schmitt-Grohé, Bob Staiger, Martín Uribe, Eric Verhoogen, Jonathan Vogel, Shang-Jin Wei, David Weinstein, and Daniel Xu; seminar participants at Columbia, Fudan, Dartmouth, Hitotsubashi, National University of Singapore, Drexel, Renmin and Tsinghua; and attendees at the Singapore Management University Urban Conference, Econometric Society China Meeting, SUFE-IESR Joint Trade Workshop, ABFER, and Urban Economic Association Meeting for helpful discussions and comments at various stages of this paper. We thank Shuting Zhang for excellent research assistance. Part of the work was done when Yang Jiao was a postdoctoral fellow at Dartmouth College, whose hospitality is greatly appreciated.

†School of Economics, Singapore Management University. Email: yangjiao@smu.edu.sg

‡INSEAD & CEPR. Email: lin.tian@insead.edu

"It's the biggest investment in high-speed internet ever. Because for today's economy to work for everyone, internet access is just as important as electricity, or water, or other basic services."

– President Joe Biden, White House address, June 2023

1 Introduction

The advent of internet has profoundly reshaped the U.S. economy. By 2013, 80% of Americans were online, a leap from the 1980s, as shown in Figure 1. The Biden-Harris Administration's recent commitment to universal high-speed internet access, under the Broadband Equity Access and Deployment (BEAD) Program, underscores the pivotal role of digital connectivity in modern economic and societal shifts.[1] Amongst its wide-ranging impacts, the increasing availability of internet has specifically reduced coordination frictions, fostering cross-regional collaborations and a more fragmented production process. Whereas extensive literature exists on geographic fragmentation across international borders through offshoring or international outsourcing (see, e.g., Hummels, Ishii and Yi, 2001; Grossman and Rossi-Hansberg, 2008), there is relatively little research on domestic production fragmentation and its impacts on the labor market.

This paper examines—theoretically, empirically and quantitatively—how internet connectivity reshapes the spatial organization of production within a country. According to a survey by the Boston Consulting Group, more than 95% of outsourcing, a key channel of production fragmentation, is conducted domestically.[2] Intuitively, the implications of this trend are distinctively different from those performed internationally due to *labor mobility*. While international borders restrict labor movement, within a country, labor can move more freely, especially in the long run, across cities. This mobility influences labor demand for various skills as economic activities shift across regions. Consequently, domestic fragmentation can lead to a redistribution of skills across local labor markets, affecting welfare, productivity, and wage inequalities both at the aggregate level and in their spatial distributions across cities.

We first develop a spatial equilibrium model in a system-of-cities setting to study how

[1]See Hjort and Tian (2024) for a survey of the recent literature.

[2]Source: BCG Global Outsourcing Survey, 2015. https://mkt-bcg-com-public-images.s3.amazonaws.com/public-pdfs/legacy-documents/file14496.pdf.

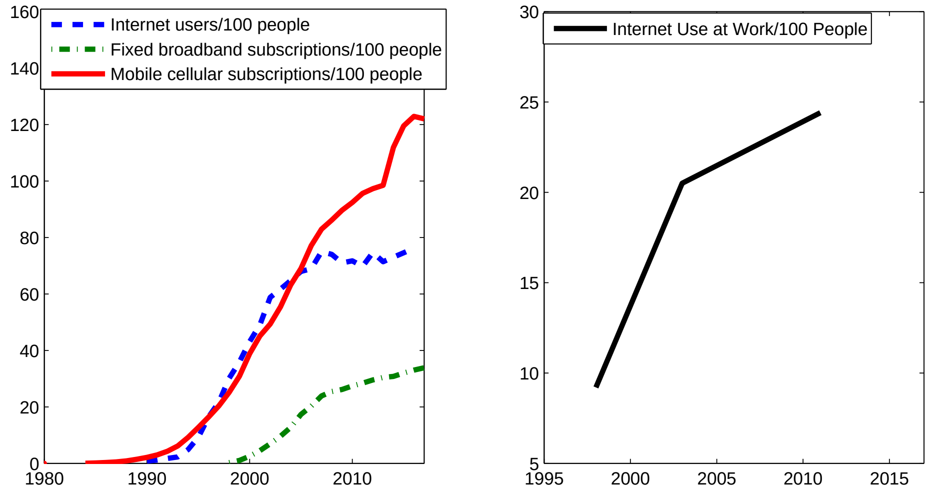


Figure 1: Increase in Internet Usage 1980 to 2013

Data source: left panel: World Development Indicators; right panel: authors' calculation from the Current Population Survey Internet and Computer Use Supplement

improvements in communication technologies facilitate the formation of cross-region production teams. In our model, the production of goods requires two distinct sets of inputs: skill-intensive knowledge inputs, produced by high-skill workers who create the "blueprint" of a product, and less skill-intensive standardized production (Garicano and Rossi-Hansberg, 2006; Arkolakis et al., 2018). Larger cities, with their comparative advantage in skill-intensive tasks (Davis and Dingel, 2019), attract a greater concentration of high-skill workers specializing in knowledge production. In contrast, workers engaged in standardized production tend to locate in smaller cities to save on costs. Both high-skill and low-skill workers are mobile across regions, with high-skill workers deciding the spatial organization of production, including location and scale, to maximize profits. Notably, the formation of cross-city production teams needs to account for fragmentation costs, such as those associated with communication and coordination when economic agents specializing in different tasks are geographically dispersed.[3] The equilibrium conditions in the model determine the extent of production fragmentation and the distribution of skills, wages, and housing prices across cities. Additionally, the model generates testable predictions on the impact of reduced communication costs on the spatial skill distribution, particularly predicting an increase in the share of high-skill workers in larger cities and a corresponding decline it in smaller cities.

[3] In this paper, we do not distinguish between "firms" and "production teams" , while allowing for collaborations in production to happen both intra- and inter-firms. For example, a furniture production team can be either an individual firm or consist of a furniture design firm and a separate furniture factory.

To validate our theoretical model, we next present empirical evidence using U.S. data. We begin by demonstrating that the observed spatial redistribution of skills over time is linked to an increasing degree of spatial segregation. Our analysis shows that larger cities have become increasingly specialized in skill-intensive activities, a trend that intensified between 1980 and 2013. During this period, high- and low-skill workers have also become more spatially segregated, consistent with the model. We further investigate the predicted impact of internet improvements on the skill composition within cities, particularly the contrasting results across cities of different sizes. To address the endogeneity of local internet quality, we adopt a novel instrumentation strategy inspired by the literature, which uses geological features for identification (e.g., Juhasz and Steinwender, 2019). Specifically, we leverage the unique features of U.S. broadband technology and use elevation levels of the local terrain to develop an instrument for internet connectivity quality. This analysis provides causal evidence supporting our main theoretical prediction, showing that improved internet connectivity increases the concentration of high-skill workers in larger cities, while reducing it in smaller cities.

Finally, we parameterize our model to quantitatively assess the effects of domestic geographic fragmentation. For this analysis, we assemble a comprehensive dataset that merges census data, internet bandwidth records, and the Orbis Database that details direct shareholder information for subsidiary plants. The Orbis data particularly enable us to measure the extent of cross-city joint production through the headquarter-subsidiary relationship. Using this dataset, we compute the bilateral fragmentation costs between city pairs and estimate key structural parameters, including the elasticity of fragmentation with respect to internet quality. We perform further quantitative analyses to evaluate the impact of internet improvements on spatial skill redistribution and the real wages of high- and low-skill workers, both directly and indirectly, through general equilibrium reallocation. In a counterfactual exercise, we simulate a world without internet improvements since the 1980s, showing that the extent of spatial skill redistribution would have been reduced by 60%, underscoring the quantitative significance of our proposed mechanism. Internet improvements have also increased real wages for both high- and low-skill workers, with a significant portion of welfare improvements driven by spatial reorganization resulting from more fragmented production across different cities. In the final exercise, we assess the policy impact of the Biden-Harris Administration's BEAD Program, showing that improvements in internet connectivity would lead to greater spatial divergence of skills, and significant welfare improvements for both high- and low-skill workers.

This paper is related to three strands of the literature. First, domestic production fragmentation is driven by similar economic forces as international production fragmentation through offshoring or outsourcing. A large volume of research studies how falling transportation or communication costs motivate firms to disintegrate production and send certain jobs overseas to take advantage of comparative advantages (see, e.g., Antràs, Garicano and Rossi-Hansberg, 2006; Grossman and Rossi-Hansberg, 2008; Abramovsky and Griffith, 2006; Fan, 2024). While the two types of production fragmentation share the same underlying driving forces, our work contrasts with this literature by focusing on the domestic context, thereby highlighting both the shared and divergent economic implications with respect to labor mobility.

Our work is also closely connected to a fast expanding literature on cross-city analysis of production fragmentation. Duranton and Puga (2005) pioneer the theoretical research, for which they develop a model with homogeneous labor that is mobile across cities and sectors, concluding that low communication cost facilitates the separation of managerial and manufacturing units in different cities. Contemporaneous works by Eckert (2019) and Eckert, Ganapati and Walsh (2022) show that ICT technology facilitates outsourcing of tradable services, which contribute to the rising wage inequality in the US. Most recently, Demir, Javorcik and Panigrahi (2023) study how fast internet access affects input sourcing and economic growth across locations, finding that firms reallocate their purchases towards suppliers with better internet and diversify their input sources. Our paper complements these studies, while contributes to the literature by demonstrating, both theoretically and empirically, the heterogeneous spatial effects of internet improvement on skill shares across cities of different sizes.

Finally, our paper is closely related to the literature on quantitative spatial equilibrium analysis, e.g., Allen and Arkolakis (2014) and Allen, Arkolakis and Takahashi (2020). Previous literature mostly focuses on transportation infrastructure, which affects trade cost. We differ by considering communication technology infrastructure, and the internet in particular. In our framework, internet improvement affects cross-city joint production cost instead of transportation cost. By doing so, our paper connects with a body of literature that studies the effect of modern technology improvement on production organization (see, e.g., Fort, 2017; Tian, 2021).

The rest of the paper is organized as follows. Sections 2 and 3 introduce the model, provide theoretical analysis, and derive equilibrium properties. Section 4 presents the empirical findings and investigates the heterogeneous effects of internet improvement on skill composition across cities of different sizes. Section 5 provides a quantitative evaluation of our model and presents results from the counterfactual exercise. Section 6 concludes.

2 The Model

Our theoretical framework embeds a model of firm production organizations in a system-of-cities setting with heterogeneous agents. The basic logic of the model can be sketched as follows: In the model, larger cities have a comparative advantage in the relatively skill-intensive managerial activities performed by high-skill workers over less skill-intensive production activities performed by lower-skill workers. A reduction in cross-city collaboration costs would induce a greater extent of cross-city collaborations (or domestic fragmentation of production), resulting in larger cities specializing more in managerial activities and smaller cities in production activities. Given spatial mobility, changes in the relative local labor demand will lead to spatial redistribution of skills, with high-skill workers becoming increasingly concentrated in larger cities.

2.1 Set-up

We consider an economy with a finite number of cities, indexed by $n \in \mathcal{N} \equiv \{1, 2, \ldots, N\}$. There is a continuum of agents, distinguished by their exogenously-given skill levels, each of whom inelastically supplies one unit of labor. The measures of high-skill workers (which we refer to as managers) and low-skill workers (which we refer to as production workers) are L^m and L^p, respectively.

Individuals consume two goods: a homogeneous tradable good and housing. The utility function follows a standard Stone-Geary form:

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

where c is the consumption of the tradable good and h is the consumption of housing. Managers and production workers choose their residential locations to maximize their utility.[4]

The homogeneous tradable good can be produced in any location of the economy with varying productivity levels. Each production team consists of a single manager and l homogeneous production workers.[5] A manager living in city n can choose to locate the production

[4]A number of papers study the various forms of mobility cost in reality; see, e.g., Moretti (2011), Baum-Snow and Pavan (2012), and Ferreira, Gyourko and Tracy (2012). This paper focuses on long-run changes in the labor market. We thus takes the position that in the long run, individuals are highly mobile.

[5]Note that this production setup is equivalent to any constant returns to scale production function. Both the high-skill input M and the low-skill input L can be equivalently translated into M production teams, each of which consists of a single manager and $\frac{L}{M}$ production workers.

team in any city $c \in \mathcal{N}$.[6] Managers living in n and producing in $c \neq n$ incur a productivity loss that we model as iceberg bilateral fragmentation costs, $\tau_{nc} \geq 1$, with τ_{nn} normalized to 1. These costs reflect the costs of managing off-site workers, e.g., communication or coordination frictions between managers and production workers located in different cities. Formally, a manager living in n and managing workers in c has the following production technology:

$$y_{nc} = \frac{a_{nc}}{\tau_{nc}} l^{\beta}. \tag{2}$$

The production technology, which follows Lucas (1978), has three elements: First, a_{nc} denotes the "manager's productivity," which we discuss further below; second, τ_{nc} reflects the iceberg productivity loss of managing an off-site production team, or fragmentation costs; and third, $\beta < 1$ is an element of diminishing returns to scale, or the manager's span of control.

A manager in city n is characterized by a productivity vector $\mathbf{a_n} = \{a_{n1}, a_{n2}, \ldots, a_{nN}\}$. These productivity vectors are origin-city specific and vary across managers, causing managers in n to potentially make different choices regarding production locations. In doing so, we assume implicitly that the productivity heterogeneities originate from managers, who take the role of developing the blueprint for the products and providing management capital for the production process.[7] The manager's productivity, a_{nc}, has two components: (1) local agglomeration force $f(L_n^m)$, which is an increasing function of the total mass of managers in city n, L_n^m; and (2) a random draw, denoted by $\bar{a}_{nc}$. The two components are assumed to enter the manager's productivity function multiplicatively:

$$a_{nc} = f(L_n^m)\bar{a}_{nc}. \tag{3}$$

In particular, a manager who lives in city n draws her productivity $\bar{a}_{nc}$ from N cities simultaneously. Each $\bar{a}_{nc}$ is drawn independently from a Fréchet distribution with a cumulative distribution function given by

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

where T_n is an exogenous technology parameter representing city n's fundamentals, such as natural resources endowment or geographic location, that potentially affect labor productivity. θ represents the dispersion of the draws: A higher value $\theta > 0$ decreases the dispersion of the manager's productivity across locations.[8]

[6] To the extent possible, we use n to denote the manager's residential location (the source of the *blueprint or management capital*) and c to index the location of the production.

[7] It is straightforward to extend the model to allow worker productivity to vary with production locations in such a way that none of the results that we focus on would be affected.

[8] The assumption of having i.i.d draws across all locations is observationally equivalent to a joint Fréchet

2.2 Manager's Optimization

In this environment, managers face a three-step optimization problem. First, a manager chooses where to live, which is also where she works and where the firm's headquarters are located. Second, the manager chooses her firm's spatial organization (i.e., location of the production team). Finally, the manager decides on the production scale (i.e., how many workers to hire). We consider the optimization problem in a backward order, starting from the last step.

2.2.1 Production Scale

Managers are the residual claimants of the firm's profit. The income of a manager who lives in city n and manages workers in city c is

$$\pi_{nc} = \frac{a_{nc}}{\tau_{nc}} l^{\beta} - w_c l, \tag{4}$$

where w_c is the wage of workers in city c. Recall also that $\tau_{nc} \geq 1$ is the iceberg cost that reflects the cost of managing workers remotely—e.g., the communication cost between city n and city c.

Given a_{nc}, a manager chooses the size of her production team, l, to maximize her income. Taking the first-order condition of (4) with respect to l, we obtain the optimal production scale l^*,

$$l^* = (\frac{\beta a_{nc}}{\tau_{nc} w_c})^{\frac{1}{1-\beta}}, \tag{5}$$

where a more productive manager (higher a_{nc}) manages a larger production team.

Combining (4) and (5), a manager living in city n with a production team in city c has an income of:

$$\pi^*_{nc} = \beta^{\frac{\beta}{1-\beta}} (1-\beta) (\frac{a_{nc}}{\tau_{nc} w_c^{\beta}})^{\frac{1}{1-\beta}}. \tag{6}$$

Note that both a higher iceberg fragmentation cost τ_{nc} and a higher worker wage w_c would reduce the manager's income.

---

distribution assumption. See Eaton and Kortum (2002), footnote 14, for a discussion.

2.2.2 Production Locations

A manager who lives in city n chooses to locate the production team in a city that maximizes her income π^*_{nc}, as specified in (6). The Fréchet assumption on the idiosyncratic component of manager's productivity allows us to derive the following "fragmentation gravity equation:"

Lemma 1 *The probability of a manager who lives in city n and locates production in city c is*

$$\frac{T_n(\tau_{nc} w_c^\beta)^{-\theta}}{\Phi_n}, \tag{7}$$

where Φ_n, city n's "fragmentation potential", is defined by

$$\Phi_n \equiv \sum_k T_n(\tau_{nk} w_k^\beta)^{-\theta}, \tag{8}$$

where $\sum_{n \in \mathcal{N}} \Phi_n = 1$.

Proof. See Appendix A. ∎

By the Weak Law of Large Numbers, the above gravity equation also gives the share of managers living in city n and locating their production teams in city c:

$$x_{nc} \equiv \frac{L^m_{nc}}{L^m_n} = \frac{T_n(\tau_{nc} w_c^\beta)^{-\theta}}{\Phi_n}. \tag{9}$$

Based on this result, it is easy to see that an internet infrastructure development that drives down cross-city fragmentation cost, τ_{nc}, increases the share of cross-city production teams relative to domestic production teams, all else equal.

2.2.3 Residential Locations

Individuals choose their residential location to maximize utility. From (1), we derive the indirect utility function for an agent with income π_n facing rent p_n in city n:

$$V(p_n, \pi_n) = \frac{\pi_n}{p_n^{1-\alpha}}. \tag{10}$$

Additionally, given the Stone-Geary preference, the equilibrium housing rent in city n is given by

$$p_n = \frac{(1-\alpha)W_n}{H_n}, \tag{11}$$

where W_n is the total income in city n, including both city n managers' and production workers' income, and H_n is the exogenously given housing supply in city n.[9]

Given the distribution of the productivity draws and the profit function π^*_{nc} in (6), we can derive the distribution for managers' income.

Lemma 2 *The income of a manager who lives in city n follows the following Fréchet distribution with a cumulative distribution function:*

$$G(\pi) = \exp\left(-[\beta^{-\beta}(1-\beta)^{-(1-\beta)}]^{-\theta} \left(f(L_n^m)\right)^{\theta} \Phi_n \pi^{-\theta(1-\beta)}\right). \tag{12}$$

Proof. See Appendix A. ■

By the properties of a Fréchet distribution, the expected income of a manager living in city n is thus

$$E[\pi_n] = \zeta[[f(L_n^m)]^{\theta}\Phi_n]^{\frac{1}{\theta(1-\beta)}}, \tag{13}$$

where $\zeta \equiv \theta\beta^{\frac{\beta}{1-\beta}}(1-\beta)^2 \int_0^{+\infty} \exp\left(-x^{-\theta(1-\beta)}\right) x^{-\theta(1-\beta)}dx$.

Managers choose their residential locations to maximize their indirect utility in (10). Denoted by Ψ_n, a manager's natural logarithm of the expected utility function is given by

$$\Psi_n = \log\left(\frac{E[\pi_n]}{p_n^{1-\alpha}}\right) = const + \frac{1}{1-\beta}\log[f(L_n^m)] + \frac{1}{\theta(1-\beta)}\log\Phi_n - (1-\alpha)\log p_n. \tag{14}$$

A manager's problem is therefore to maximize Ψ_n. In a spatial equilibrium, managers are indifferent between living in city n and n' (conditional on there being non-zero managers in both cities), so that $\Psi_n = \Psi_{n'}$, $\forall n, n' \in \mathcal{N}$, or

$$\begin{aligned} &\frac{1}{1-\beta}\log[f(L_n^m)] + \frac{1}{\theta(1-\beta)}\log\Phi_n - (1-\alpha)\log p_n \\ =&\frac{1}{1-\beta}\log[f(L_{n'}^m)] + \frac{1}{\theta(1-\beta)}\log\Phi_{n'} - (1-\alpha)\log p_{n'}. \end{aligned} \tag{15}$$

[9] In the baseline model, we assume that the housing supply is fixed. In Appendix K, we relax this assumption and consider a scenario in which housing supply is elastic.

2.3 Worker's Optimization

Similar to managers, production workers also choose their location to maximize their indirect utility in (10), given their income w_n and housing price p_n. In equilibrium, ex-ante homogeneous production workers are indifferent and thus receive the same indirect utility across cities, i.e., $V_n^w = V_{n'}^w = \bar{v}\ \forall\ n,\ n' \in \mathcal{N}$. We therefore obtain the following equilibrium condition:

$$w_n/p_n^{1-\alpha} = w_{n'}/p_{n'}^{1-\alpha}. \tag{16}$$

3 Equilibrium Analysis

In this section, we characterize the spatial equilibrium. We first provide the definition, then consider a special "fragmentation autarky" case to provide more intuition. We next focus on a simplified two-city model to derive analytic results for the effects of changes in fragmentation costs, τ_{nc}, on the distribution of skills. We finally perform numerical simulation in a multi-city scenario.

3.1 Definition

In a spatial equilibrium, managers and workers are indifferent across locations.[10] With exogenous parameters $\{T_n, \tau_{nc}, H_n\}\,\forall n, c \in \mathcal{N}$, and a mass of managers L^M and workers L^P, an equilibrium is a vector of labor allocations $\{L_{nc}^m, L_{nc}^P\}$ and prices $\{p_n, w_n\}$ such that:

1. Production workers maximize their utility in (10);
2. Housing prices p_n are determined by (11);
3. Managers maximize their expected utility in (14);
4. Labor markets clear for both managers and workers:

$$L^m = \sum_n L_n^m = \sum_{n,c} L_{nc}^m, \tag{17}$$

[10]Given the unbounded Frechét distribution draws, we can show that all cities have a nonzero mass of production workers. Additionally, to be consistent with the data, we further assume that city fundamentals ensure that all cities have a nonzero mass of managers.

and

$$L^p = \sum_n L_n^p = \sum_{n,c} L_{nc}^p, \tag{18}$$

where L_{nc}^p refers to the mass of production workers hired by managers from n and living in city c. This is given by

$$L_{nc}^p = \eta w_c^{-1} \left(T_n (\tau_{nc} w_c^{\beta})^{-\theta} \right) \Phi_n^{\frac{1}{\theta(1-\beta)}-1} [f(L_n^m)]^{\frac{1}{1-\beta}} L_n^m, \tag{19}$$

where $\eta = \beta^{\frac{1}{1-\beta}} \int_0^{\infty} y^{-\frac{1}{\theta(1-\beta)}} e^{-y} dy$.[11]

In Appendix C, we provide further details on the equilibrium characterization. Furthermore, we provide, using Banach fixed point theorem, a set of sufficient conditions under which the equilibrium exists and is unique.

3.2 Equilibrium with Infinite Fragmentation Cost

To derive the analytic results, we adopt the following parametric assumption for the agglomeration force,

$$f(L_n^m) = (L_n^m)^{\gamma},$$

where the $\gamma \geq 0$ parameter governs the strength of agglomeration externalities.

We first consider a special "fragmentation autarky" case, in which the bilateral fragmentation cost $\tau_{nc} \to +\infty$, $\forall n \neq c$. In this scenario, the system of equilibrium conditions reads as follows:

$$\gamma(\log L_n^m - \log L_{n'}^m) = (\log w_n - \log w_{n'}) - (\log T_n^{\frac{1}{\theta}} - \log T_{n'}^{\frac{1}{\theta}}) \tag{20}$$

and

$$\left(\frac{1}{1-\alpha} + \frac{\beta}{1-\beta} \right) (\log w_n - \log w_{n'}) = \left(\frac{\gamma}{1-\beta} + 1 \right) [\log L_n^m - \log L_{n'}^m] + \frac{1}{1-\beta} [\log T_n^{\frac{1}{\theta}} - \log T_{n'}^{\frac{1}{\theta}}]. \tag{21}$$

In this case, the cross-city fragmentation cost is prohibitively high, such that all managers will hire production workers in the same city as the one in which the manager lives. We can show that under regularity conditions, i.e., $\gamma + 1 > \frac{\gamma}{1-\alpha}$, cities with high technology parameters (T_n) not only have a larger fraction of the whole population, but also have a

[11] See Appendix D for details on derivation of the demand for production workers.

larger fraction of both the high-skill population and the low-skill population.[12] Formally, we state the results in the following proposition.

Proposition 3 *Given $f(L_n^m) = (L_n^m)^\gamma$ and $\gamma + 1 > \frac{\gamma}{1-\alpha}$, 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 satisfy that*

$$L_n^m \propto T_n^\kappa, \tag{22}$$

$$L_n^p \propto T_n^\kappa, \tag{23}$$

where $\kappa = \frac{\frac{1}{1-\alpha}-1}{1+\gamma-\frac{\gamma}{1-\alpha}}\frac{1}{\theta} > 0$. As a result, the high-skill employment share $\frac{L_n^m}{L_n^m+L_n^p}$ is the same across all cities.

Proof. See Appendix A. ■

3.3 Equilibrium with Finite Fragmentation Costs

We next analyze the equilibrium with finite fragmentation costs. We start with a simple two-city case to elucidate the mechanism of skill relocation after a reduction in fragmentation costs, before extending the analysis to a multi-city scenario.

3.4 Two-city Analysis

We start with a two-city case with quasi-symmetric communication cost and fixed housing supply.[13] Using the simple model, we highlight the mechanism behind the skill relocation after cross-city communication cost reduction.

First, it is easy to see that when the cross-city fragmentation cost is infinite, the city with the greater technology parameter is larger and more skill intensive. We can solve for $\frac{L_1^m}{L_2^m}$ and

[12] The assumption that $\gamma + 1 > \frac{\gamma}{1-\alpha}$ implies that the elasticity of agglomeration, which is positively correlated with γ, is smaller than the elasticity of urban costs, which is positively correlated with $1-\alpha$. This ensures that cities have a finite size in equilibrium. See Behrens, Duranton and Robert-Nicoud (2014) for related discussions.

[13] We assume $H_n = H_c = 1$ in this section to highlight the role of the comparative advantage of cities in technology T_n. The analysis can easily be extended to the case with different city-level housing supplies. In our quantitative section, we take into account housing supply heterogeneity.

$\frac{w_1}{w_2}$ explicitly using (20) and (21):

$$\log \frac{L_1^m}{L_2^m} = \frac{\frac{\alpha}{1-\alpha}}{\gamma + 1 - \frac{\gamma}{1-\alpha}} [\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] \tag{24}$$

and

$$\log \frac{w_1}{w_2} = \frac{1}{\gamma + 1 - \frac{\gamma}{1-\alpha}} [\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}]. \tag{25}$$

Suppose, without loss of generality, that $T_1 > T_2$. From (24) and (25), it is obvious that the population in city 1 is higher than that in city 2. With a reduction in communication cost, we can show that a small reduction in fragmentation cost—e.g., an improvement in internet quality that facilitates cross-city communication—results in a spatial reallocation of skills. Specifically, the share of high-skill employment in the initially larger city will increase, whereas the share of high-skill employment in the initially smaller city will decrease.

Proposition 4 *In the two-city case, if $\tau_{12} = \tau_{21}$ goes down around the neighborhood of the infinite communication cost, and suppose that $T_1 > T_2$, then L_1^m and L_2^p would go up, L_2^m and L_1^p would go down. A stronger agglomeration force (larger γ) implies larger labor reallocation for both the high-skilled and the low-skilled.*

Proof. See Appendix A. ■

This proposition states that if internet improvement reduces cross-city communication cost, then bigger cities will attract a larger proportion of the high skilled as internet quality improves. Through a numerical simulation of the two-city equilibrium, we confirm the proposition's prediction on skill flows after the internet improvement. As shown in the left panel of Figure 2, when the ICT openness—defined as the inverse of the fragmentation costs—increases, the share of managers increases in the larger city, whereas the share of managers goes down in the smaller city, as shown in the right panel of Figure 2.[14]

We also consider the welfare implications of a reduction in communication cost. In general, simulations support the notion that both managers and workers benefit from communication cost reduction. Intuitively, the drop in the iceberg communication cost is similar to the productivity increase in the production function in a Hick's neutral way. However, deriving an analytic result for the welfare impact is hard with the spatial reorganization, because the real wages of managers and workers are both endogenous and depend on each other.[15] We

[14] Formally, ICT openness, denoted by $\triangle$, is defined as $\tau^{-\theta}$.

[15] This is because both managers and workers consume housing, and housing prices enter into the welfare functions of both types of agents.

provide local analyses of agents' welfare in Appendix B. In our quantitative analysis section, we directly examine the welfare implications with calibrated parameters.

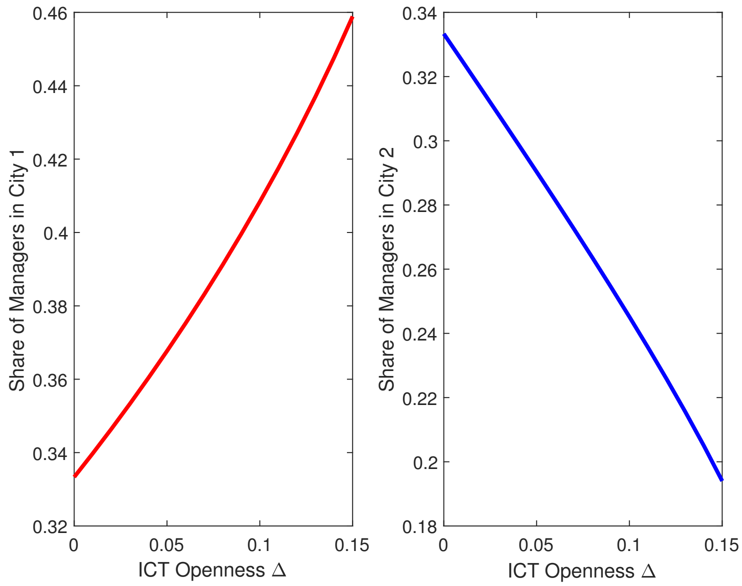


Figure 2: Two-city Equilibrium: High-skill Employment Share in City 1 and City 2

Notes: In this simulation, we set $\theta = 5$, $T_1^{\frac{1}{\theta}} = 1.5$, $T_2^{\frac{1}{\theta}} = 1.0$, $L_m = 1$, $L_p = 2$, $\gamma = 0.1$, $\alpha = 0.4$, $\beta = 0.4$, $H_1 = H_2 = 1.0$. The vertical axes show the share of managers in City 1 (the larger city) in the left panel and that in City 2 (the smaller city) in the right panel. The horizontal axis shows the extent of ICT openness $\triangle \equiv \tau^{-\theta}$.

3.5 Multi-city Analysis

We now turn to a multi-city analysis to explore the heterogeneous effects of a reduction in the fragmentation cost on city-level skill composition. The objective is to have a relatively large number of cities to mimic the fact that the population of every city, even the largest city, constitutes only a minor fraction of the total population, so a single city's internet improvement would not affect the other cities significantly. At the same time, we want to avoid simulating too many cities, which is computationally intensive but does not provide additional insights in a qualitative fashion. To this end, we choose an eight-city scenario, in which we consider the case in which all cities have the same housing supply and there are four big cities and four small cities with technology parameters given as follows:

$$T_1 = T_2 = T_3 = T_4 > T_5 = T_6 = T_7 = T_8.$$

The results are displayed in Figure 3. Figure 3a shows that if there is an internet improvement in a small city, say City 8, which reduces the bilateral fragmentation costs between the city and all the other seven cities, then the share of managers decreases in City 8. The intuition is that as a small city, the low-skill wage here is relatively lower. When the city gets more connected with the rest of the nation, some managers in those bigger cities find it more profitable to relocate their production teams in the smaller city, which increases the local demand for production workers there, giving rise to an inflow of the low-skill workers in the smaller city. In contrast, Figure 3b shows that if there is an internet improvement in a big city, say City 4, which reduces the bilateral fragmentation costs between this city and all the other seven cities, then the share of managers increases in this city. The intuition is that as a big city, the low-skill wage here is relatively higher. When the city gets more connected with the rest of the nation, some managers from smaller cities find it more profitable to relocate to that city themselves to leverage the strong agglomeration externalities there, while keeping their production teams in other low-cost small cities. In doing so, larger cities attract an inflow of high-skill labors.

These sets of results, together with the numerical simulations in the two-city case, illustrate our key theoretical result stated in Proposition 4. The logic behind this result can be found in a standard Ricardian model. When there is a drop in communication cost—i.e., trade cost—different regions specialize in the activities in which they have a comparative advantage. For larger cities, these are skill-intensive management-related tasks, whereas in small cities, this corresponds to less skill-intensive standardized production. Given that factors are mobile within a country, changes to local labor demand driven by specialization will imply a redistribution of high- and low- skill workers across space, with larger cities receiving an inflow of high-skill workers and smaller cities an inflow of low-skill workers.

In summary, the model developed not only illustrates the key mechanisms of geographic production fragmentation, but also generates the following testable predictions:

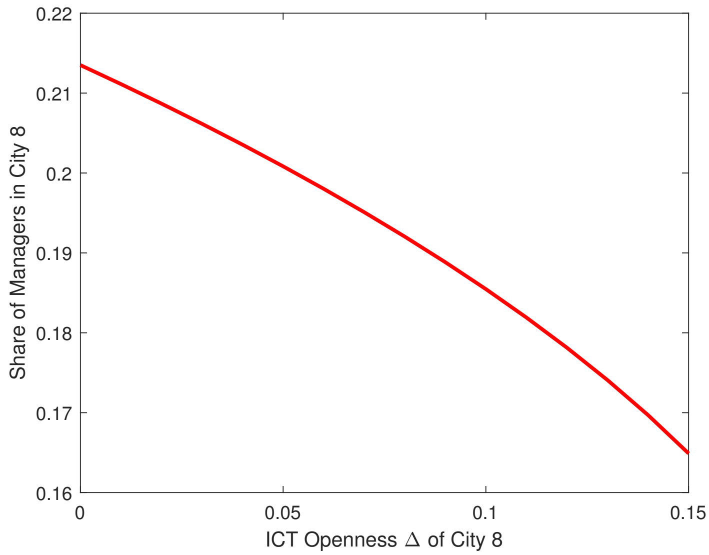
1. Larger cities specialize in skill-intensive activities;

2. Overtime (as internet improves), the extent of spatial segregation of skills increases; and

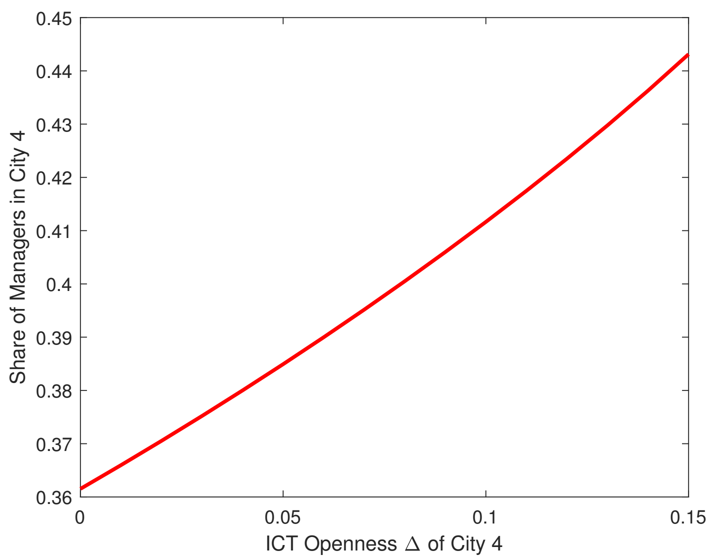
3. Local internet improvement drives up the high-skill employment share in bigger cities and reduces it in smaller cities.

In the next section, we provide empirical validation of the model predictions.

Figure 3: Eight-City Equilibrium



(a) Eight-City Equilibrium: Share of Managers



(b) Eight-City Equilibrium: Share of Managers

Notes: Figures 3a and 3b display the results of an eight-city simulation. In this simulation, we set $\theta = 5$, $T_1^{\frac{1}{\theta}} = T_2^{\frac{1}{\theta}} = T_3^{\frac{1}{\theta}} = T_4^{\frac{1}{\theta}} = 1.5$, $T_5^{\frac{1}{\theta}} = T_6^{\frac{1}{\theta}} = T_7^{\frac{1}{\theta}} = T_8^{\frac{1}{\theta}} = 1.0$, $L_m = 1$, $L_p = 2$, $\gamma = 0.1$, $\alpha = 0.4$, $\beta = 0.4$, $H_1 = H_2 = H_3 = H_4 = H_5 = H_6 = H_7 = H_8 = 1.0$. The vertical axes show the share of managers in City 8 (the smaller city) in the top panel and that in City 4 (the larger city) in the bottom panel. The horizontal axis shows the extent of ICT openness $\triangle \equiv \tau^{-\theta}$.

4 Data and Stylized Facts

In this section, we examine the model predictions using data on U.S. cities. We first document stylized facts that establish the the relationship between spatial skill redistribution and the trend of rising production fragmentation: (1) larger cities have become increasingly specialized in skill-intensive activities; and (2) between 1980 and 2013, there had been a substantial increase in the extent of spatial segregation of skills across U.S. cities. Next, we validate the key theoretical result stated in Proposition 4 by examining the heterogeneous effects of fragmentation cost reduction on the share of high-skill employment across cities of different sizes.

4.1 Data Description

Our analysis mainly draws on the Integrated Public Use Micro Samples (IPUMS, Ruggles et al., 2015). For 1980, we use 5% Census samples; for later years, we combine the 2011, 2012, and 2013 1% American Community Survey (ACS) samples. Our worker sample consists of individuals who were between the ages 16 and 64 and who were working in the year preceding the survey. Residents of institutional group quarters, such as prisons and psychiatric institutions, are dropped along with unpaid family workers.

We define a city as a commuting zone (CZ), which is the geographic unit of analysis developed by Tolbert and Sizer (1996). Each CZ is a cluster of counties characterized by strong commuting ties within and weak commuting ties across zones. For our analysis, we include 722 CZs in the continental U.S. We measure city size using the log of labor supply, which is measured by the product of weeks worked times the usual number of hours worked per week.[16] All calculations are weighted by the Census sampling weight multiplied by the labor supply weight.

Throughout the paper, we classify workers into high- and low-skill groups using their occupation wage in 1980. Following Acemoglu and Autor (2011), we rank the skill levels of different occupations, approximated by the mean log hourly wage of workers in each occupation in 1980.[17] We define high-skill workers as those whose occupation wage rank is

[16] We use labor supply instead of number of workers to measure city population to be consistent with our use of hourly wage. The stylized facts are robust when we use number of workers, and are available upon request.

[17] Examples of occupations in the lower, middle, and upper wage-rank distributions are child-care workers, waiters and waitresses, housekeepers, and hotel clerks; machine operators, reception and information desk, typists, and carpenters; CEOs, engineers, architects, financial managers, and software developers, respectively.

higher than 75% of occupations in 1980. We vary the cutoff in robustness checks to 67% and 80%. Further robustness checks using education information to classify the high and low skilled, with the high skilled defined as those with a college education or above, are also provided in the Appendix.

The internet quality data is drawn from the U.S. Federal Communications Commission (FCC) Fixed Broadband Deployment Database. Fixed broadband providers are required to provide the lists of census blocks in which they offer service in at least one location within the block. The database, available from December 2014, also provides additional information about the quality of the service, including download and upload bandwidths (reported in megabytes per second). We identify the maximum bandwidth at the block level and compute the population-weighted internet quality at the CZ level.[18]

Finally, in the quantitative exercise in Section 5, we measure the extent of firm fragmentation using the Orbis Database for 2018 from Bureau van Dijk, which reports direct shareholder information for subsidiary plants. We define a headquarters-subsidiary pair if a headquarters has strictly more than 50% of the ownership of a given subsidiary. Using location information, we construct a CZ-pair-level fragmentation measure that counts the number of headquarters-subsidiary pairs for a given pair of CZs.

4.2 Stylized Facts on Skill Distribution

Using the IPUMS data set, we document patterns of spatial skill redistribution across U.S. cities and investigate how the spatial pattern is related to domestic production fragmentation.

The left panel of Figure 4 depicts a well-known fact: The largest cities, measured by the labor supply, are the ones that have the highest shares of high-skill employment in both 1980 and 2013. This pattern of skill specialization suggests comparative advantage differences across cities of different sizes: Larger cities have a comparative advantage in more skill-intensive activities, possibly due to stronger agglomeration forces, and smaller cities have a comparative advantage in less skill-intensive activities, aided by the lower labor costs.

Moreover, this pattern of specialization became more pronounced over time. The right panel of Figure 4 plots the change in the skilled share within a city between 1980 and 2013 against the corresponding city size in 1980.[19] It shows that larger cities experienced a greater

[18]The 15-digit census block ID comes from the 2010 census. In computing the population-weighted average internet quality measures, we use the 2010 population information at the PUMA level—the smallest geographic unit in the 2010 Census. We aggregate data from the more finely divided census block level to the PUMA level using simple averages.

[19]Figure 12 in the Appendix provides the scatter plot for the raw data.

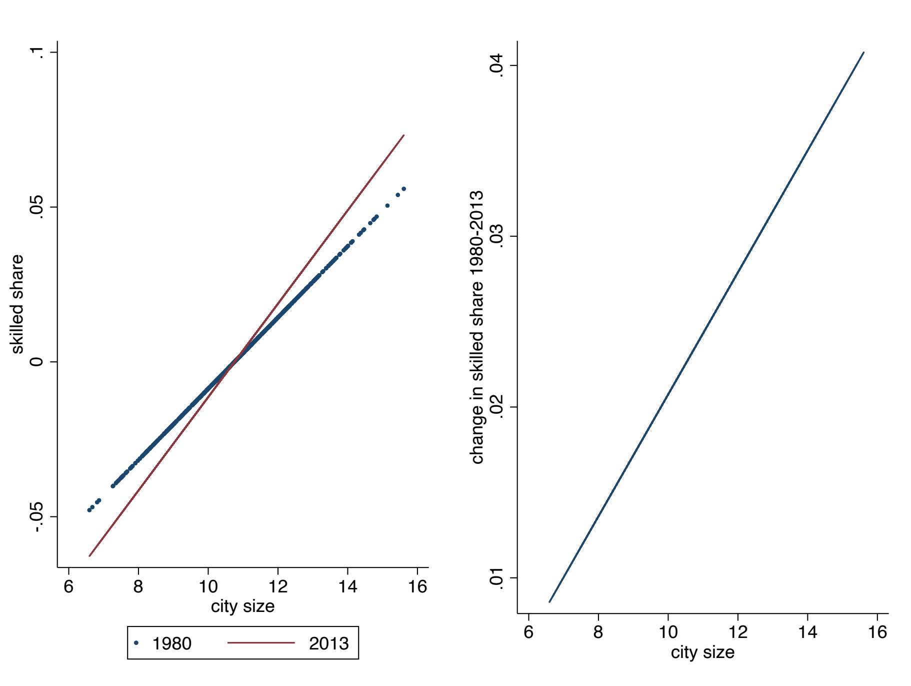


Figure 4: Change in High-skill Employment Share with Respect to City Sizes

Notes: The left panel displays the regression line for the high-skill share (demeaned) in 1980 and 2013 against log of 1980 labor supply. The right panel displays the change in the skilled share from 1980 to 2013. High skill is defined as occupation rank above 75% using the 1980 mean of log hourly wage.

increase in the share of high-skill employment between 1980 and 2013, thereby becoming even more skill intensive. For example, the share of high-skill employment in the largest city in the U.S. had risen by 4 percentage points, whereas it only increased by less than 1 percentage point in the bottom percentile of the city-size distribution. Table 1 reports the formal statistical test, in which we regress changes in the share of high-skill employment at the city level onto the city sizes measured by the log of 1980 city-level labor supply. We find strong positive correlation between city size and magnitude of the change in the high-skill employment share.[20]

Dependent variable:	Change in high-skill employment share 1980-2013	
	(1)	(2)
City Size	0.0036***	0.0050***
	(0.001)	(0.001)
State fixed effect	No	Yes
Observations	722	722
R^2	0.037	0.357

$* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$

Table 1: Change in High-skill Employment Share and City Size

Notes: City size is measured by the log of total labor supply in 1980 within a commuting zone. High-skill workers are defined as workers whose occupation wage rank is higher than 75% of occupations in 1980. Column (1) reports results using robust standard errors, and Column (2) reports results with standard errors clustered by state.

The spatial redistribution of skills, established in the previous set of results, suggests an increase in the spatial segregation of high- and low-skill workers over the period between 1980 and 2010. To study this spatial segregation more directly, we adopt a variant of the Kremer and Maskin (1996) measure of the degree of segregation, i.e.,

$$\rho = \frac{1}{S} \sum_{s} \left[\frac{\sum_c N_{cs} \cdot (\pi_{cs} - \pi_s)^2}{N_s \cdot \pi_s \cdot (1 - \pi_s)} \right],$$

where $s \in \{1, 2, \ldots, S\}$ denotes a sector as defined by Census *ind1990* codes, N_{cs} is the employment in sector s and city c, N_s is the total sectoral employment, $\pi_{cs} = \frac{N_{cs}^{skilled}}{N_{cs}}$ is the high-skill employment share in sector s and city c, and $\pi_s = \frac{N_s^{skilled}}{N_s}$ is the high-skill

[20] Table 14 in the Appendix reports robustness checks regarding the definition of the high skilled using a 67% occupation wage rank cutoff and an 80% occupation wage rank cutoff, as well as an analysis that uses education to separate high- and low-skill workers.

employment share in sector s.[21] As shown in Table 2, the Kremer and Maskin (KM) index, denoted by ρ, almost tripled from 1980 to 2013, indicating that the high skilled and low skilled had become increasingly more spatially segregated.

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

Table 2: KM Segregation Index in 1980 and 2013

Notes: High-skill workers are defined as workers whose occupation wage rank is higher than 75% of occupations in 1980. The 95% confidence interval of the index of segregation is:

$$\frac{F(N-J,J-1)_{0.025}}{F(N-J,J-1)_{0.025}+\frac{1-\rho}{\rho}} \leq \tilde{\rho} \leq \frac{F(N-J,J-1)_{0.975}}{F(N-J,J-1)_{0.975}+\frac{1-\rho}{\rho}},$$

where $J = C + S$ (Kremer and Maskin, 1996).

In summary, we establish that larger cities have a comparative advantage in skill-intensive activities (Fact 1). This pattern of specialization became stronger over the past three decades, as high- and low-skill workers become more segregated geographically (Fact 2). In Appendix E, we provide further empirical evidence that links the observed increase in spatial segregation with our proposed mechanism of increasing production fragmentation across U.S. cities. In particular, we show that this pattern of segregation across space at the industry level is closely related to production fragmentation activities in the U.S. economy.

4.3 Heterogeneous Effects of Internet Quality on City Skill Composition

To validate the key theoretical result stated in Proposition 4, we next empirically investigate the heterogeneous effects of fragmentation cost reduction on the share of high-skill employment across cities of different sizes, i.e., local internet improvement drives up the high-skill employment share in bigger cities and reduces it in smaller cities.

[21] This index measures how correlated the employment shares of different occupations are within a city-sector. It is constructed as the ratio of the variance of share of the high-skill across cities to the variance of an agent's occupation status (i.e. the high skilled vs. the low skilled) of the a given sector, which is equivalent to the R^2 value of a regression of the share of high skilled on a series of city dummies. When $\rho = 0$, there is no segregation; i.e. the high skilled and low skilled are always in the same cities; when $\rho = 1$, there is complete spatial segregation of the high skilled and low skilled. We calculate the national average as an average value across sectors to account for possible changes in industry composition within cities across time.

4.3.1 Empirical Specification

We examine these predictions by presenting evidence from U.S. cities. We employ a long difference exercise with the following specification:

$$\Delta_t l_n = \beta_1 + \beta_2 \Delta_t q_n + \beta_3 \left(L_{n,t_0} \times \Delta_t q_n\right) + \gamma \mathbf{X_{n,t_0}} + \epsilon_n, \tag{26}$$

where $\Delta_t l_n \equiv \Delta_t \left(\frac{L_n^m}{L_n}\right)$ is the change in high-skill employment share in city n between 1980 and 2013, L_{n,t_0} is the log of total labor supply in city n in 1980, $\Delta_t q_n$ is the change in internet quality in city n between 1980 and 2013, and $\mathbf{X_{n,t_0}}$ is other controls including state dummies and initial city size.

Our key parameters of interest are the coefficients on internet quality and the interaction term between city size and internet quality, i.e., β_2 and β_3. The model predicts that $\beta_2 < 0$ and $\beta_3 > 0$, which imply that internet quality improvement in a small city will reduce the skilled employment share locally, while internet quality improvement in a big city will increase the local skilled employment share.

As explained earlier, local internet quality is measured using FCC internet infrastructure data. Specifically, we first calculate the simple average of upload and download bandwidths at the PUMA level, denoted by b_{nj}^{up} and b_{nj}^{down} in PUMA j of CZ n. The CZ-level internet quality measure, q_n, is defined as

$$q_n = \sum_{j \in G(n)} \frac{\log(1 + b_{nj}^{up}) + \log(1 + b_{nj}^{down})}{2} \frac{pop_{nj}}{\sum_{k \in G(n) pop_{nk}}}, \tag{27}$$

where $G(n)$ is the set of PUMAs in CZ n and pop_{nj} is the population in PUMA j of CZ n.[22] Given that there is virtually no commercial use of the internet in 1980, q_n in 2013 also represents the change in internet quality from 1980 to 2013, i.e., $\Delta_t q_n \equiv q_n$. Figure 5 shows a map of internet qualities across CZs. Notably, there is large variation in internet quality across different regions.

4.3.2 Internet Quality, City Size, and the Skilled Employment Share

Using the internet data, we run the specification in (26). The first two columns of Table 4 report the results. Columns (1) and (2) show the OLS results without and with state fixed effects. Importantly, consistent with the model predictions, we find that $\hat{\beta}_2 < 0$ and $\hat{\beta}_3 > 0$.

[22] Taking *log* reduces potential outliers within a CZ. Adding 1 to the measured bandwidths ensures that when there is no internet available, $b_{nj}^{up} = 0$ and $b_{nj}^{down} = 0$, we have $q_n = 0$ as well.

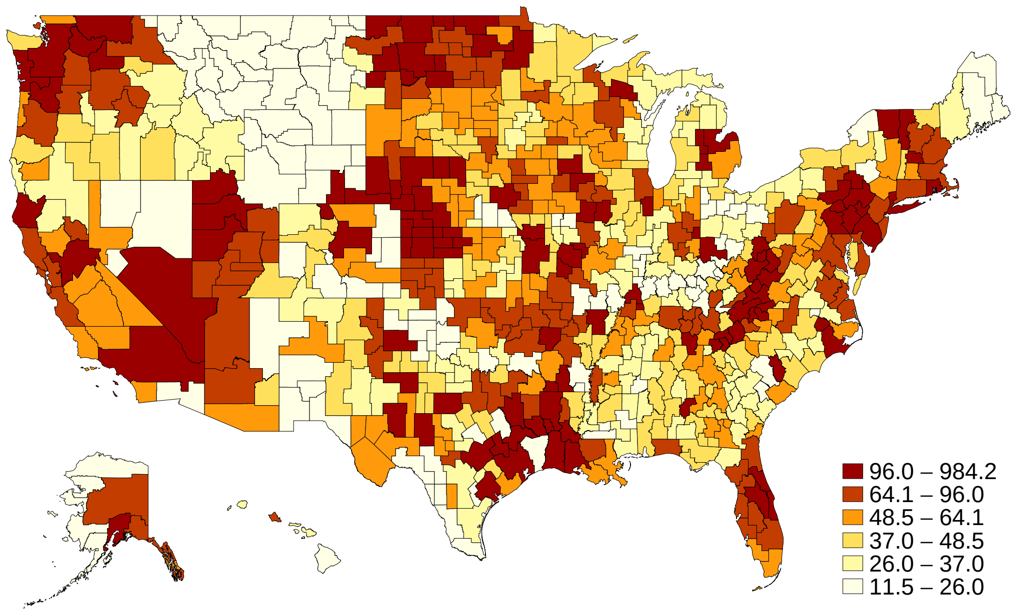


Figure 5: Average Internet Bandwidth in U.S. Commuting Zones

Notes: This figure displays the average internet bandwidth in the U.S. at CZ level, calculated as the population-weighted average of upload and download bandwidths. Bandwidths are measured in megabytes per second.

Both are statistically significant at the 5% level. These two results jointly imply that better internet quality reduces a small city's high-skill employment share and increases a big city's high-skill employment share, thereby confirming our model's predictions.

An obvious problem that arises when estimating (26) using OLS is that internet quality is endogenous. Specifically, there are three major concerns: First, there may be long-run local employment trends that drive both the internet provision and high-skill employment share. Second, there may be unobserved local shocks that affect both internet quality improvement and changes in high-skill share over time. The third concern is reverse causality: Local demand shocks for skills may drive internet provision. It is worth noting, however, that for both the second and third points above, the potential bias must work in opposite directions for larger vs. smaller cities to generate results consistent with theoretical predictions. For example, for the reserve causality to work here, one must form a theory whereby local internet quality improvement is driven by a *larger* share of high-skill workers in *bigger* cities but a *lower* share of high-skill workers in *smaller* cities. Similarly, for the second concern, the omitted variable must both increase internet quality and *high*-skill share in *bigger* cities, while increasing internet quality and *low*-skill share in *smaller* cities.

To address the first concern, in Columns (4) and (5) of Table 4, we perform a falsification test by replacing the left-hand-side variable with the change in the high-skill employment share from an earlier period, between 1950 and 1980. The estimates show that there is indeed no role for later development of the internet in the change in skilled employment share in earlier years, thus ruling out the existence of long-run local employment trends.

To address the second and third concerns, we consider an instrumental variable approach, using topographic elevation and initial telecommunication infrastructure before 1980 as the exogenous determinant in the provision of broadband internet. The use of geographic elements, such as topography, as a means for identification frameworks is common in many empirical studies because of their generally random and predetermined nature, as we see in, e.g., Miguel, Satyanath and Sergenti (2004) and Juhasz and Steinwender (2019). We follow this general approach and develop an instrument for broadband internet in the U.S. using terrain elevations. The instrument leverages a unique feature of the US broadband internet provision, in which, unlike many other developed countries, the most used technology for signaling distribution relies on cable infrastructure. The key intuition is that low-lying areas are more prone to floods and exhibit higher summer temperatures, and such climatic conditions play a crucial role in driving up the costs of deployment and maintenance of cable broadband infrastructure, thereby leading to worse internet qualities in these areas (Jaber, 2013; Amorim, Lima and Sampaio, 2022). To formally establish the predictive power of the instrument on internet provision, we use the following "stage-zero" analysis:

$$q_n = \alpha_1 + \alpha_2 elevation_n + \alpha_3 ini\_telephone\_penetration_n + \mathbf{X_n} + \epsilon_n, \tag{28}$$

where q_n is the population-weighted log of average upload and download speeds in CZ n defined in (27), $Elevation_n$ is the population-weighted average terrain elevation in CZ n, and X_n includes city size and dummy variables for state fixed effects. Additionally, we control for initial telephone penetration in the CZ, $ini\_telephone\_penetration_n$, which is calculated as the average fraction of households that have access to telephones in 1970 and 1980 within CZ n. In Table 3, we find that $\hat{\alpha}_2 = 0.193$ and $\hat{\alpha}_3 = 2.849$, both of which are statistically significant at the 5% level. This shows that all else equal, places with higher terrain tend to receive better internet, lending support to the relevance assumption of the instrument.

In addition to the relevance condition, the instrument must also satisfy the following exclusion restriction: Terrain level must affect the change in skill share only through its impact on internet quality. This is likely to hold for two reasons. First, while terrain is likely correlated with other factors that may affect the *level* in the share of high-skill workers through other channels, it is unlikely that they affect the *flow* of high-skill workers between the two periods. For this to happen, the correlation between the instruments and these other

Dependent variable: Internet quality		
	(1)	(2)
elevation	0.214**	0.193**
	(0.085)	(0.080)
initial telephone penetration		2.849***
		(0.656)
State Fixed Effect	Y	Y
Observations	722	722
R^2	0.442	0.461

Table 3: Terrain Elevation, Initial Telephone Penetration and Internet Quality

Notes: This table reports the relationship between internet quality (year 2014) and terrain elevation and initial telephone penetration (before 1980). Robust standard errors are in parentheses. City size and state fixed effects are controls.

factors will have to become stronger over time. Next, the correlations will also have to be systematically *different* across city sizes, i.e., the correlations are positive for larger cities (thereby attracting an inflow of high-skill workers) and negative for smaller cities (thereby resulting in an outflow of high-skill workers).

Results from the 2SLS estimation, which are shown in Column (3) of Table 4 remain qualitatively consistent from the OLS estimates. The point estimates of the coefficients on q_n and $L_{n,t_0} \times q_n$ are greater in value than the OLS estimates. This may be due to classical measurement errors in the regressors, which would result in attenuation bias. Crucially, the model predicts that $\hat{\beta}_2 < 0$ and $\hat{\beta}_3 > 0$ continue to hold. In Column (6), we further apply the 2SLS to our placebo test for the 1950-1980 change in the share of high-skill employment. We draw a similar conclusion there is no long-run local employment trend that drives our results on internet quality and high-skill employment share change.

Dependent variable: Change in the share of high-skill employment						
	1980-2013			1950-1980		
	OLS		2SLS	OLS		2SLS
Estimates	(1)	(2)	(3)	(4)	(5)	(6)
q_n	-.023**	-.029**	-.137***	-.005	-.012	-.013
	(.009)	(.012)	(.034)	(.017)	(.020)	(.037)
$L_{n,t_0} \times q_n$	.0022**	.0028**	.011***	-0.000	.001	.001
	(.0008)	(.0011)	(.003)	(.001)	(.001)	(.003)
State Fixed Effects	No	Yes	Yes	No	Yes	Yes
Observations	722	722	722	722	722	722
R^2	.045	.360	-0.076	.048	.284	-0.338
S-W F-stats (First Stage)						
Internet quality			12.92			12.92
Internet quality × city size			11.15			11.15

Table 4: Heterogeneous Effects of Internet Improvement on Skill Shares Across Cities

Notes: City size is measured by log(labor supply 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 are clustered at the state level when there are state fixed effects. We also report Sanderson-Windmeijer (S-W) F-statistics for the first-stage regressions. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$

5 Quantitative Analysis

Empirical validation of key model predictions lends credibility to our theoretical framework. We next carry out quantitative analysis using the model. We first calibrate the model parameters, then conduct two counterfactual exercises by changing internet qualities in the US. For both exercises, we consider the consequences of the counterfactual changes in internet qualities on spatial skill distributions and real wages of high- and low-skill workers.

5.1 Parametric assumptions

For the quantitative assessment, we maintain the functional form assumption for the agglomeration force: $f(L_n^m) = (L_n^m)^\gamma$, where $\gamma \geq 0$ is the parameter measuring the extent of regional agglomeration. Moreover, we parameterize the bilateral fragmentation cost as follows:

$$\log \tau_{nc} = \lambda_{nc} + \delta^d \log d_{nc} + \delta^I q_{nc}. \tag{29}$$

Equation (29) assumes that the fragmentation cost between two cities n and c takes a semi-parametric form, i.e., a power function of the bilateral geographic distance d_{nc} between the two cities and the quality of the internet connection between the two cities q_{nc}, in addition to a term λ_{nc} that summarizes all other associated costs—e.g., if the two cities are located in the same state, and if the two cities share a common border. We refer to δ^d as the distance elasticity of joint production and δ^I as the internet elasticity of joint production.

For the bilateral internet connection, we assume it adopts a quasi-symmetric form such that

$$q_{nc} = q_n \times q_c, \tag{30}$$

where q_n is city n's internet quality defined in (27). Note that by using the interaction term $q_n \times q_c$, we allow potential complementarity in both cities' internet quality. For instance, if there is no internet in city c, then the bilateral communication cost between c and n will remain very high, regardless of n's internet quality.

One concern is that internet quality may also change trade cost in goods, which can possibly change the skill distribution as well. In Appendix F, we empirically evaluate the role of internet quality in goods trade using the commodity flow survey data. We show that the internet does not have any significant impact on bilateral goods trade. Admittedly, the internet has certainly changed the economy in many other ways in addition to reducing the fragmentation cost. However, focusing on the fragmentation cost is useful for understanding the impact of the internet through this specific channel.

5.2 Calibration of Parameters

In this section, we calibrate model parameters. We begin by assigning values to parameters that have been estimated in past literature. We then describe in detail the estimation procedures for the other parameters, including agglomeration externalities, dispersion of manager productivity, city-specific housing supply, and technology parameter, as well as the fragmentation costs.

Parameters from Previous Literature

For some parameters in our model that are commonly used in the literature, we adopt their values directly. Specifically, we use existing estimates for the values of the share of spending on housing, $1-\alpha$, and the span of control, β. We set $1-\alpha$ at 0.24 (Davis and Ortalo-Magné, 2011; Behrens, Duranton and Robert-Nicoud, 2014) and β at 0.53 (Buera and Shin, 2013).

See Table 5 for details.

Parameter	Value	Description	Moment / Source
$1-\alpha$	0.24	Share of spending on housing	Literature
β	0.53	Span of control	Literature
γ	0.010	Agglomeration externality	Elasticity of worker wage wrt city size
θ	4.11	Frechét dispersion parameter	High-skill workers income distribution
δ^d	0.230	Distance elasticity of joint productions	Implied from gravity estimates
δ^I	-0.010	Internet elasticity of joint productions	Implied from gravity estimates

Table 5: Calibrated Model Parameters

Calibration of γ

The strength of agglomeration forces γ is set to target the elasticity of average worker wage with respect to city size. The implied value of γ is 0.010, which is broadly in line with the agglomeration externalities estimated in past literature (see, e.g., Combes and Gobillon, 2015). In Appendix H, we conduct a sensitivity analysis to show that changing the value of γ does not impact the results in a significant manner.[23]

Calibration of θ

The Frechét distribution parameter θ determines the dispersion of managers' income across cities. From the cumulative distribution function of manager's income in (12), we obtain

$$-\log[-\log G(\pi)] = \theta(1-\beta)\log \pi + \log[(L_n^m)^{\gamma\theta}\Phi_n] + constant. \tag{31}$$

We use the 3-year ACS 2011-2013 data to obtain the high-skilled hourly wage distribution. Using the individual hourly wage information, we run an OLS regression with city fixed

[23]Note that γ is distinct from the observed productivity advantages of cities. In the model, larger cities are more productive for three reasons: (1) the "raw" agglomeration externalities captured by γ; (2) the exogenous productivity differences across cities implied by T_n; and (3) the sorting of high- and low-skill workers because of fragmentation. Furthermore, γ is a term that summarizes multiple aspects of skill-biased agglomeration externalities, including productivity, housing and amenities (Diamond, 2016).

effects, which absorbs the $\log[(L_n^m)^{\gamma\theta}\Phi_n]$ term. The OLS estimation then gives $\theta(1-\beta) = 1.93$, which implies a value of 4.11 for θ.

Calibration of Fragmentation Cost

Recall that the bilateral fragmentation cost is assumed to take the following functional form:

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

The key parameter of interest is δ^I, the elasticity of fragmentation with respect to internet quality. The estimation difficulty is that the aggregate cross-city fragmentation cost τ_{nc} is not directly observed. To overcome this, we rely on the gravity equation derived in (9). We first compute X_{nc}, the number of occurrences of the joint productions in city c that originate from city n, by multiplying both sides of (9) by the total number of managers in origin city L_n^m:

$$X_{nc} = L_n^m \frac{T_n \tau_{nc}^{-\theta} w_c^{-\beta\theta}}{\Phi_n}. \tag{32}$$

Normalizing by X_{nn} and using the assumption that $\tau_{nn} = 1$, we obtain the following function that links τ_{nc} with city-level worker wages and X_{nc}:

$$\tau_{nc} = \left(\frac{w_c^{\beta\theta} X_{nc}}{w_n^{\beta\theta} X_{nn}} \right)^{-1/\theta}. \tag{33}$$

Since w_n is directly observed our data set, we can calculate τ_{nc} using additional information on cross-city joint productions, i.e., X_{nc}. We rely on multi-locational production data to measure X_{nc}. The data are constructed using the Orbis Database, which reports ownership information for subsidiary plants. We define a headquarters-subsidiary pair if a headquarters has strictly more than 50% of the ownership of a given subsidiary. Moreover, the database reports the locations of the subsidiary and the headquarters, which allows us to count the number of headquarters-subsidiary pairs at the city-pair level. Specifically, for each city c, we calculate X_{nc} by counting the number of subsidiaries that belong to headquarters located in a given commuting zone n.

Admittedly, these headquarters-subsidiary pairs by no means capture all of the cross-city joint-production forms; e.g., firm's domestic outsourcing is not included. However, given this data limitation, we view that this headquarters-subsidiary pair as a reasonable starting point to study this question, for two reasons: First, the headquarters-subsidiary relationship fits the high-skill and low-skill joint production setting well in the theoretical part; and second, it identifies a specific channel through which firms can achieve fragmented production.

Using the estimates of τ_{nc} from (33), we can estimate (29) using the following specification:

$$\log \tau_{nc} = \chi_n + \iota_c + \delta^d \log d_{nc} + \delta^I q_{nc} + \Theta \mathbf{H_{nc}} + \varepsilon_{nc}, \tag{34}$$

where χ_n and ι_c are origin and destination fixed effects, respectively, d_{nc} is the distance between two cities n and c, q_{nc} denotes the bilateral internet connectivity between city n and city c as defined in (30), and $\mathbf{H_{nc}}$ is a vector of city-pair controls, including state pair fixed effects, the interaction between city sizes, dummies for two cities sharing a border, dissimilarities in the language spoken, and racial mix.[24]

We estimate Equation (34), where coefficient estimates $\hat{\delta}^d$ and $\hat{\delta}^I$ are reported in Table 6. We find that as expected, greater geographical distance reduces the number of headquarters-subsidiary pairs. More importantly, the result shows that a higher quality of bilateral internet connectivity induces more headquarters-subsidiary pairs. Additional sensitivity analysis by changing the values of δ^I is presented in Appendix I.

Dependent variable: Fragmentation cost between n and c (τ_{nc})		
Estimates	(1)	(2)
$\log d_{nc}$	0.283***	0.134***
	(0.002)	(0.004)
q_{nc}	-0.048***	-0.010***
	(0.003)	(0.003)
Controls	No	Yes
City Fixed Effects	Yes	Yes
N	44,188	44,023

Table 6: Gravity Equation Estimates

Notes: Robust standard errors in parentheses. Significance levels: * 10%, ** 5%, ***1%. Controls include a dummy on whether two CZs share a border, dissimilarity in language spoken, dissimilarity in race mix, interaction between both two CZs' size and state pair fixed effects.

[24] The dissimilarity in language spoken is constructed as follows: For each commuting zone, we use 1980 census data to calculate the fractions of people that speak English, Spanish, French, German, and other languages at home. Then for any two commuting zones, we compute the Euclidean distance of the fractions. The dissimilarity in race is constructed using four racial categories—white, black, native, and others—and follow the same approach as in computing the dissimilarity in language spoken.

Housing Supply H_n

Another set of parameters in the model is the exogenous housing supply, which is estimated using ACS 2011-2013 data on the average city-level wage w_n for workers and total labor income W_n of the city, i.e.,

$$\frac{H_n}{H_{n'}} = \frac{W_n / w_n^{\frac{1}{(1-\alpha)}}}{W_{n'} / w_{n'}^{\frac{1}{(1-\alpha)}}}. \tag{35}$$

We normalize the housing supply in CZ with code 00100 as 1—i.e., $H_{CZ00100} = 1$—and then compute H_n for other cities using the ratio above.

City Technology T_n

The final step requires estimating the city-specific technology parameter T_n. Combining the definition for the city-level fragmentation potential Φ_n in (8) and the equilibrium condition for manager's living location choice

$$\frac{\gamma}{1-\beta} \log \frac{L_n^m}{L_c^m} + \frac{1}{\theta(1-\beta)} \log \frac{\Phi_n}{\Phi_c} = (1-\alpha) \log \frac{p_n}{p_c} = \log \frac{w_n}{w_c},$$

we obtain:

$$\frac{\gamma}{1-\beta} \log \frac{L_n^m}{L_c^m} + \frac{1}{\theta(1-\beta)} \log \frac{T_n}{T_c} + \frac{1}{\theta(1-\beta)} \log \frac{\sum_k (\tau_{nk} w_k^\beta)^{-\theta}}{\sum_k (\tau_{ck} w_c^\beta)^{-\theta}} = \log \frac{w_n}{w_c}. \tag{36}$$

Given the set of $\{L_n^m,\ \tau_{nc}^{-\theta},\ w_n\}$, we can back out T_n. We normalize $T_{CZ00100} = 1$ and then pick T_n so that the model-implied values for $\log \frac{T_n}{T_c}$ match the corresponding values estimated. Figure 6 shows the model-generated technology parameters. We can see that CZs with greater technology parameters are concentrated in large cities (e.g., New York, San Francisco, and Seattle) and other denser areas along the coasts. Regressing the calibrated values of $log(T_n)$ on city sizes yields a positive and statistically significant coefficient, with point estimate at 0.0595 and a 95% confidence interval at [0.0378, 0.0811].

Finally, we calibrate our model parameters without targeting the number of workers in each city directly. However, from the model, in each city, the total labor income equals the sum of the number of workers of each type multiplied by the average income of each type. As an external validation test, we check whether our calibrated model delivers a good match of the number of low-skill labor (thus the total number of workers) in each city. Figure 13 in Appendix N confirms that this is indeed the case.

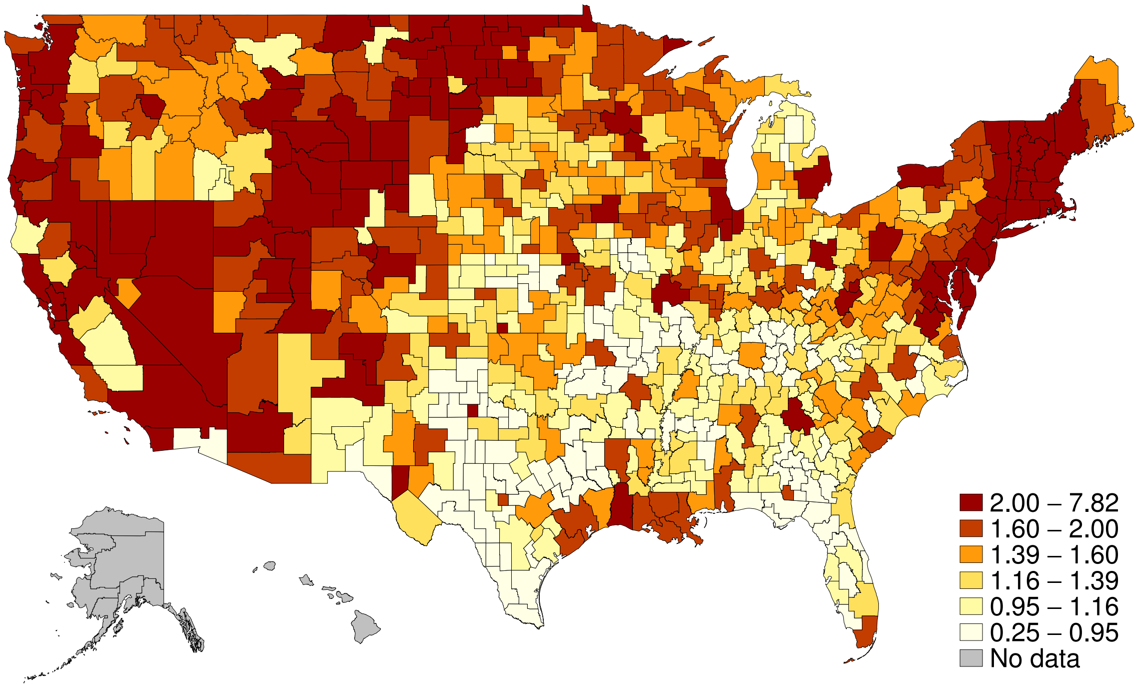


Figure 6: City-specific Technology Parameters

Notes: The figure shows the model-generated technology parameter (T_n) in 2013.

5.3 Internet Infrastructure, Skill Relocation, and Welfare

We perform an exercise in which we assume there is no internet quality improvement between 1980 and 2013. This counterfactual scenario allows us to evaluate the role of internet infrastructure in the skill relocation in the U.S. through production fragmentation, and assess the impact of internet improvement on welfare of high- and low-skill workers.

Specifically, we first take the calibrated parameters in 2013 as given. We then solve the model under two different sets of bilateral fragmentation costs. The first set of fragmentation costs is directly obtained from the data in 2013 using equation (33). The second set of fragmentation costs is derived by assuming a counterfactual scenario in which the cost $\tilde{\tau}_{nc}$ is otherwise identical to the estimated cost except for internet connectivity—i.e., $\tilde{\tau}_{nc} = \exp(\log(\tau_{nc} - \hat{\delta}^I q_{nc}))$. The model is solved numerically under these two scenarios for 722 cities, using a global method. Appendix G provides details of the numerical methodology.

We then compare the high-skill share in each city under the two scenarios. The reduction in the share of high-skill workers informs us of the contribution of internet technology in driving the observed changes in skill concentration across space. Figure 7 visualizes the positive relationship between the change in high-skill employment share and city size. In the left panel, both the baseline scenario (with internet) and the counterfactual scenario (without internet) exhibit a positive slope. The flatter slope under the counterfactual scenario shows that the extent of skill redistribution would have been smaller without internet quality improvement. Formally, we regress the change in high-skill share on city size and obtain a coefficient of 0.0031 when state fixed effects are included, as shown in Table 7. Comparing this against the observed skill redistribution reported in Table 1, it implies that without internet connectivity, the observed skill redistribution in the US would have been reduced by about $0.0030/0.00503 = 60\%$.

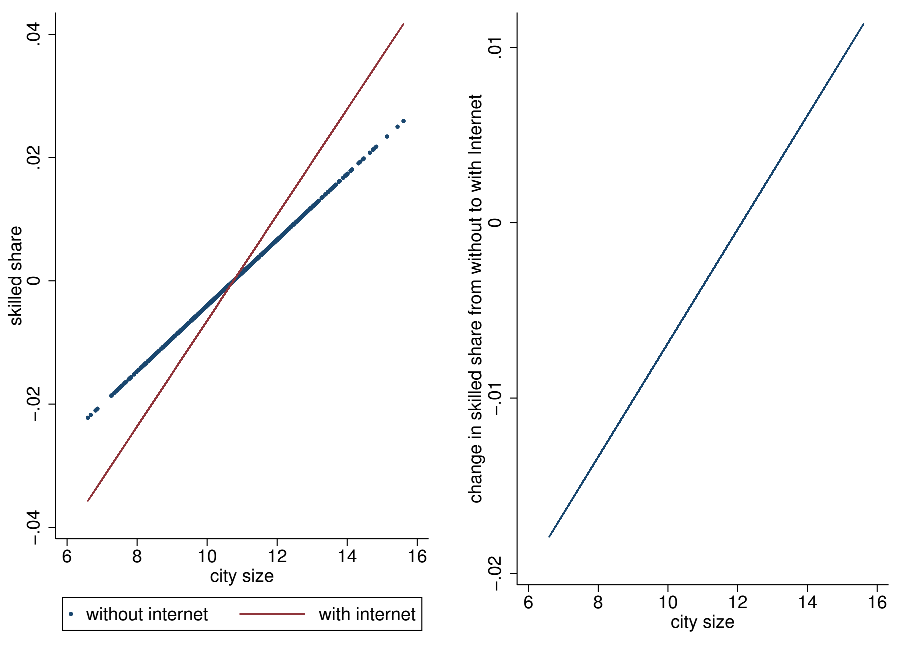


Figure 7: Simulated Relationship between High-skill Employment Share and City Size

Notes: This figure displays the model generated change in high-skill employment share when moving the economy from the case without internet and with internet in 2013. City size is measured as log(labor supply in 1980), as in regression Table 1.

We also study the welfare implications of internet infrastructure through the lens of production fragmentation and find that its impact is sizable. With the internet, production workers' welfare (real consumption) increases by 3.62% and managers' welfare by 3.39%. The intuition is that the reduction in fragmentation cost is effectively technological progress from the whole economy's point of view. Moreover, the reduction in fragmentation cost does

Dependent variable:	Change in high-skill employment share with internet	
	(1)	(2)
City Size	0.0031***	0.0030***
	(0.0006)	(0.0009)
State fixed effects	No	Yes
Observations	722	722
R^2	0.050	0.132

$*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$

Table 7: Model-implied Change in High-skill Employment Share and City Size

Notes: The dependent variable is the change in high-skill employment share when moving the economy from the case without the internet to the case with the internet. City size is measured by log(labor supply in 1980). Column (1) reports results using robust standard errors, and Column (2) with standard errors clustered by state.

not exhibit skill bias and benefits the joint production of managers and production workers. Therefore, it increases the welfare of managers and production workers in similar magnitudes. Additionally, internet technological progress, as shown above, drives spatial reorganization of production, by changing local demand for and high- and low-skill workers across cities of different sizes. This spatial reorganization benefits workers of both skill types as well.

The welfare implications from internet improvement can be attributed to two channels: the direct effect driven by the drop in the iceberg fragmentation cost τ, and the indirect general equilibrium effect from spatial reorganization. We further decompose the welfare changes into these two components. To get the first component, we fix all the production teams in the scenario without the internet. Suppose that there are L^m_{nc} managers from city n cooperate with L^p_{nc} production workers from city c to produce, calculated in the scenario without the internet. The total output is also a function of the bilateral fragmentation cost τ_{nc} since it works as an iceberg cost in our model. We denote their output as $\frac{1}{\tau_{nc}} S_{nc}(L^m_{nc}, L^p_{nc})$, where S_{nc} is the aggregate production function at the bilateral level (see Appendix J for the detailed derivations of function S_{nc}). Since in each team, managers get $1-\beta$ share of the output and production workers get β share of the output. We then compute the housing price in any city n as the total income in city n divided by housing supply in city n

$$p^x_n = \frac{(1-\beta)\sum_c \frac{1}{\tau^x_{nc}} S_{nc}(L^m_{nc}, L^p_{nc}) + \beta \sum_c \frac{1}{\tau^x_{cn}} S_{cn}(L^m_{cn}, L^p_{cn})}{H_n}, \tag{37}$$

where x denote the two scenarios, "with internet" and "without internet." We then compute the log-change in the real income of managers and production workers under the two sets of fragmentation cost $\tau_{nc}^{with\ internet}$ and $\tau_{nc}^{no\ internet}$. For managers, the change in real income directly due to the drop in fragmentation cost is

$$\log\left[(1-\beta)\sum_{n,c}\frac{1}{\tau_{nc}^{with\ internet}}\frac{S_{nc}(L_{nc}^{m},L_{nc}^{p})}{(p_n^{with\ internet})^{1-\alpha}}\right]-\log\left[(1-\beta)\sum_{n,c}\frac{1}{\tau_{nc}^{no\ internet}}\frac{S_{nc}(L_{nc}^{m},L_{nc}^{p})}{(p_n^{no\ internet})^{1-\alpha}}\right]. \tag{38}$$

For production workers, the change in real income directly due to the drop in fragmentation cost is

$$\log\left[\beta\sum_{n,c}\frac{1}{\tau_{cn}^{with\ internet}}\frac{S_{cn}(L_{cn}^{m},L_{cn}^{p})}{(p_n^{with\ internet})^{1-\alpha}}\right]-\log\left[\beta\sum_{n,c}\frac{1}{\tau_{cn}^{no\ internet}}\frac{S_{cn}(L_{cn}^{m},L_{cn}^{p})}{(p_n^{no\ internet})^{1-\alpha}}\right]. \tag{39}$$

The difference between the total changes in welfare and the above changes—which is the direct welfare implications from internet improvement—gives the welfare change component due to spatial reorganization for managers and production workers, respectively. Table 8 reports the decomposition of welfare changes. While the direct effect of the drop in the fragmentation cost accounts for the major increase in both managers' and workers' real incomes, the general equilibrium effect of spatial reorganization on welfare is also not negligible. It accounts for more than one seventh of the welfare increase for managers and about one fifth of the welfare increase for production workers.

	Δ Managers' Welfare	Δ Workers' Welfare
Directly Due to Fragmentation Cost Change	2.89%	2.93%
Due to Spatial Reorganization	0.50%	0.69%
Total	3.39%	3.62%

Table 8: A Decomposition of Welfare Change

Notes: This table shows the total welfare (real income) increase and decompose it to two components. The first component is directly due to the change in fragmentation cost. The second component is the remaining part and labeled as "due to spatial reorganization".

5.4 Narrowing the Digital Divide

In the final analysis, we conduct a policy experiment to assess the potential effects of improving internet access and quality in underdeveloped areas, a common objective for many governments and international organizations aiming to bridge the "digital divide." For example, the BEAD Program plans to invest over $40 billion to provide reliable and affordable high-speed internet to both residential and business locations in the US. Given the significant disparities in internet quality across U.S. cities, we are interested in understanding the potential impact of improving internet quality in currently underserved areas on skill redistribution and overall welfare. To examine this, we consider an experiment that focuses on enhancing the internet speed for underserved business units in each city—specifically those with speeds below 100/25 Mbps, in line with the BEAD program's objectives. We hypothesize a scenario where the government allocates $20 billion (approximately half of the BEAD funding) to upgrade the internet speed of these underserved business locations to 300/300 Mbps.[25]

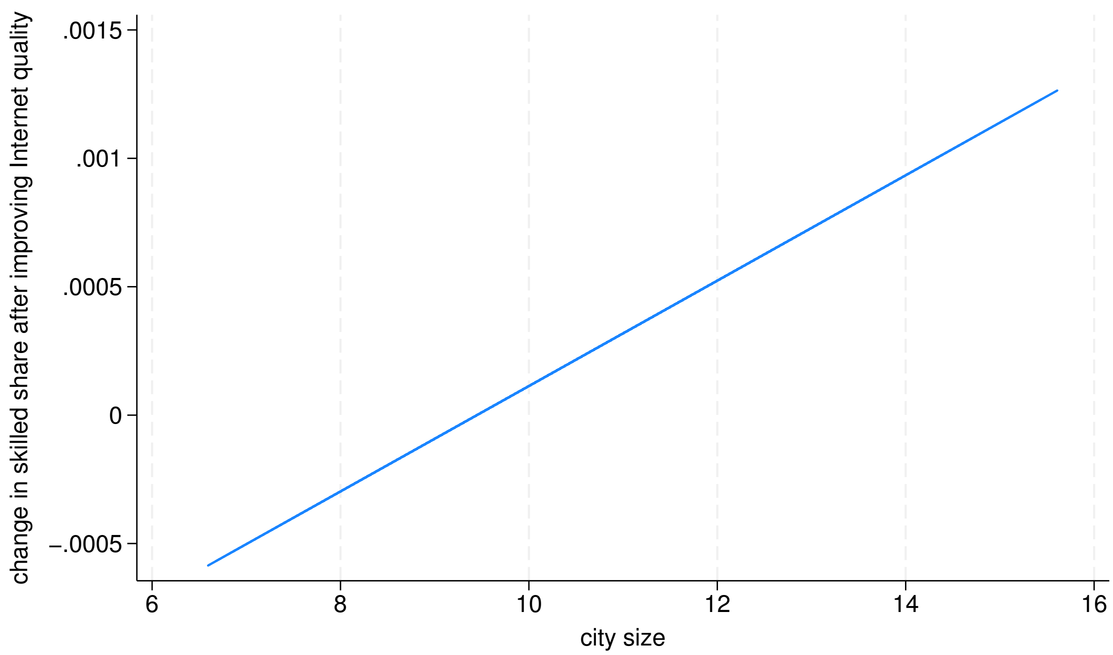


Figure 8: Change in High-skill Share after Improving Internet Quality

Notes: This figure displays the regression line of the change in the skilled share against city size after improving underserved locations' internet speed in U.S. The estimated slope is 0.0085 with a robust standard error 0.0014.

Figure 8 shows the relationship between the model-generated change in the high-skilled workforce share and city size following the simulated improvements in internet quality. The positive correlation indicates that, on average, larger cities will attract a relatively greater proportion of high-skilled workers, thereby intensifying the spatial divergence of skills across cities. The welfare implications of this shift are significant: our findings show that the welfare

[25]We compiled a dataset detailing the total number of business locations and the number of underserved business locations in each city. Assuming that the government spends $30 per month (a figure close to the fees charged by U.S. internet providers) to improve internet speed at one underserved business location over a 5-year period, we calculate the proportion of underserved business locations that could be covered by this funding. This allows us to estimate the average increase in internet quality across cities.

of production workers would increase by 0.27%, while managers' welfare would also rise by 0.27%. This welfare improvement is partly driven by the direct benefits of enhanced internet quality and partly by the general equilibrium effect—whereby managers reorganize their production teams across cities to capitalize on the comparative advantages of different locations. These results suggest that improving internet quality in poorly connected areas can yield substantial benefits. This analysis provides a quantitative assessment of internet infrastructure investments, offering valuable insights for cost-benefit analyses of policies aimed at reducing the digital divide.

6 Conclusion

This paper examines the impact of reduced communication costs on geographic fragmentation of domestic production. Our spatial equilibrium model, which incorporates cross-city production and skill heterogeneity, demonstrates that advancements in communication technologies facilitate geographic fragmentation, leading to significant changes in the spatial organization of production. These changes, in turn, drive the concentration of high-skill workers in larger cities and reduce their presence in smaller cities, contributing to a new pattern of skill distribution.

Our empirical findings, validated through a novel instrumental variable approach, confirm that improved internet connectivity plays a pivotal role in these spatial redistributions. Quantitatively, we show that internet improvements account for a significant share of the observed changes in skill distribution from 1980 to 2013, highlighting the critical role of reduced communication costs in shaping modern labor markets. These changes have broader welfare implications, arising from both direct effects of reducing fragmentatio costs and the indirect general equilibrium effect of spatial reorganization. While our paper abstracts from the complexities of firm boundaries and simplifies the production organization structure to two layers, future research could benefit from explicitly considering the role of firm boundaries and the multi-layered hierarchy of production in geographic fragmentation.

References

Abramovsky, Laura, and Rachel Griffith. 2006. "Outsourcing and Offshoring of Business Services: How Important is ICT?" *Journal of the European Economic Association*, 4(2/3): 594–601.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*. Vol. 4, , ed. Orley Ashenfelter and David Card, Chapter 12, 1043–1171. Elsevier.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics*, 129(3): 1085–1140.

Allen, Treb, Costas Arkolakis, and Yuta Takahashi. 2020. "Universal Gravity." *Journal of Political Economy*, 128(2): 393–433.

Amorim, Guilherme, Rafael Costa Lima, and Breno Sampaio. 2022. "Broadband Internet and Protests: Evidence from the Occupy Movement." *Information Economics and Policy*, 60: 100982.

Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg. 2006. "Offshoring in a Knowledge Economy." *Quarterly Journal of Economics*, 121(1): 31–77.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple. 2018. "Innovation and Production in the Global Economy." *American Economic Review*, 108(8): 2128–73.

Baum-Snow, Nathaniel, and Ronni Pavan. 2012. "Understanding the City Size Wage Gap." *Review of Economic Studies*, 79(1): 88–127.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy*, 122(3): 507 – 553.

Buera, Francisco J., and Yongseok Shin. 2013. "Financial Frictions and the Persistence of History: A Quantitative Exploration." *Journal of Political Economy*, 121(2): 221–272.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics*. Vol. 5, , ed. Gilles Duranton, J. Vernon Henderson and William C. Strange, 247–348. Elsevier.

Davis, Donald R., and Jonathan I. Dingel. 2019. "A Spatial Knowledge Economy." *American Economic Review*, 109(1): 153–70.

Davis, Morris A., and François Ortalo-Magné. 2011. "Household Expenditures, Wages, Rents." *Review of Economic Dynamics*, 14(2): 248–261.

Demir, Banu, Beata Javorcik, and Piyush Panigrahi. 2023. "Breaking Invisible Barriers: Does Fast Internet Improve Access to Input Markets?" Working Paper.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *American Economic Review*, 106(3): 479–524.

Duranton, Gilles, and Diego Puga. 2005. "From Sectoral to Functional Urban Specialisation." *Journal of Urban Economics*, 57(2): 343–370.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741–1779.

Eckert, Fabian. 2019. "Growing Apart: Tradable Services and the Fragmentation of the U.S. Economy." Society for Economic Dynamics 2019 Meeting Papers 307.

Eckert, Fabian, Sharat Ganapati, and Conor Walsh. 2022. "Urban-Biased Growth: A Macroeconomic Analysis." National Bureau of Economic Research Working Paper 30515.

Fan, Jingting. 2024. "Talent, Geography, and Offshore R&D." *The Review of Economic Studies*, rdae044.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy. 2012. "Housing Busts and Household Mobility: An Update." *Economic Policy Review*, 18(3).

Fort, Teresa C. 2017. "Technology and Production Fragmentation: Domestic versus Foreign Sourcing." *Review of Economic Studies*, 84(2): 650–687.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics*, 121(4): 1383–1435.

Giannone, Elisa. 2022. "Skill-Biased Technical Change and Regional Convergence."

Grossman, Gene M., and Esteban Rossi-Hansberg. 2008. "Trading Tasks: A Simple Theory of Offshoring." *American Economic Review*, 98(5): 1978–97.

Hjort, Jonas, and Lin Tian. 2024. "The Economic Impact of Internet Connectivity in Developing Countries."

Hummels, David, Jun Ishii, and Kei-Mu Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics*, 54(1): 75–96.

Jaber, Ahmed. 2013. "Broadband Internet and Political Behavior: Evidence from the United States."

Juhasz, Réka, and Claudia Steinwender. 2019. "Spinning the Web: Codifiability, Information Frictions and Trade."

Kremer, Michael, and Eric Maskin. 1996. "Wage Inequality and Segregation by Skill." National Bureau of Economic Research Working Paper 5718.

Lucas, Robert. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics*, 9(2): 508–523.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy*, 112(4): 725–753.

Moretti, Enrico. 2011. "Local Labor Markets." In *Handbook of Labor Economics.* Vol. 4, , ed. Orley Ashenfelter and David Card, Chapter 14, 1237–1313. Elsevier.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. "Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]." *Minneapolis: University of Minnesota.*

Tian, Lin. 2021. "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil." Centre for Economic Policy Research Discussion Paper 16590.

Tolbert, Charles M., and Molly Sizer. 1996. "US Commuting Zones and Labor Market Areas: A 1990 Update." United States Department of Agriculture, Economic Research Service Staff Reports 278812.

Online Appendix

A Proofs

This section provides proofs to results presented in Section 2.

A.1 Proof of Proposition 1

Proof. Denote $X_{nc} = \frac{\bar{a}_{nc}}{\tau_{nc} w_c^{\beta}}$, then

$$G_{nc}(x) = Pr(X_{nc} \leq x) = Pr(\bar{a}_{nc} \leq \tau_{nc} w_c^{\beta} x) = e^{-T_n(\tau_{nc} w_c^{\beta})^{-\theta} x^{-\theta}}.$$

Define

$$X = \max_c X_{nc}.$$

Then

$$G_n(x) = Pr(X \leq x) = \Pi_{c=1}^{N} G_{nc}(x) = e^{-\Phi_n x^{-\theta}}.$$

The probability that city c provides the highest x to n is:

$$Pr[X_{nc} \geq \max\{x_{ns}; s \neq c\}] = \int_0^{\infty} \Pi_{s \neq c}[G_{ns}(x)] dG_{nc}(x) = \frac{T_n(\tau_{nc} w_c^{\beta})^{-\theta}}{\Phi_n},$$

where $\Phi_n \equiv \sum_k T_n(\tau_{nk} w_k^{\beta})^{-\theta}$. ■

A.2 Proof of Proposition 2

Proof.

$$\begin{aligned}
Pr(\pi_n \leq k) &= Pr\left[\beta^{\frac{\beta}{1-\beta}}(1-\beta)[f(L_n^m)]^{\frac{1}{1-\beta}} \max_c\{(\frac{\bar{a}_{nc}}{\tau_{nc} w_c^{\beta}})^{\frac{1}{1-\beta}}\} \leq k\right] \\
&= Pr\left[\max_c \bar{a}_{nc} \leq \beta^{-\beta}(1-\beta)^{-(1-\beta)} \tau_{nc} w_c^{\beta} k^{1-\beta} / [f(L_n^m)]\right] \\
&= e^{-[f(L_n^m)]^{\theta} \Phi_n [\beta^{-\beta}(1-\beta)^{-(1-\beta)}]^{-\theta} k^{-\theta(1-\beta)}}.
\end{aligned}$$

■

A.3 Proof of Proposition 3

Combining equations (24) and (25) by eliminating $\log w_n - \log w_{n'}$, we obtain that

$$[1+\gamma-\frac{\gamma}{1-\alpha}][\log L_n^m - \log L_{n'}^m] = [\frac{1}{1-\alpha}-1]\frac{1}{\theta}\log\frac{T_n}{T_{n'}}.$$

Therefore,

$$\log L_n^m - \log L_{n'}^m = \kappa[\log T_n - \log T_{n'}], \tag{40}$$

where

$$\kappa = \frac{\frac{1}{1-\alpha}-1}{1+\gamma-\frac{\gamma}{1-\alpha}}\frac{1}{\theta} > 0,$$

given that $1+\gamma > \frac{\gamma}{1-\alpha}$.

That is,

$$\frac{L_n^m}{L_{n'}^m} = \left(\frac{T_n}{T_{n'}}\right)^{\kappa}$$

Then we get that

$$L_n^m \propto T_n^{\kappa}. \tag{41}$$

From equation (24), we have

$$\log w_n - \log w_{n'} = \gamma[\log L_n^m - \log L_{n'}^m] + \frac{1}{\theta}[\log T_n - \log T_{n'}] = [\gamma\kappa + \frac{1}{\theta}][\log T_n - \log T_{n'}].$$

And from equation (19), we have workers' mass in city n is given by

$$L_n^p \propto w_n^{-1} T_n w_n^{-\beta\theta} T_n^{\frac{1}{\theta(1-\beta)}-1} w_n^{-\beta\theta[\frac{1}{\theta(1-\beta)}-1]} (L_n^m)^{\frac{\gamma}{1-\beta}+1}.$$

We can combine the above three equations to arrive at

$$L_n^p \propto T_n^{\kappa}. \tag{42}$$

It is then also clear that given a set of $\{T_n\}$, the equilibrium exists and is unique.

The skill premium is read directly from equation (56). When $\tau_{nc} \to \infty$, $\forall\ n \neq c$, it is written as

$$\frac{\gamma}{1-\beta}\log L_n^m + \frac{1}{1-\beta}\log T_n^{\frac{1}{\theta}} - \frac{1}{1-\beta}\log w_n.$$

From equation (19), we get

$$\log w_n = \log T_n^{\frac{1}{\theta}} + (\gamma + 1 - \beta) \log L_n^m - (1-\beta) \log L_n^p + (1-\beta) \log \eta.$$

The skill premium is then equal to a constant. Note that η is only a function of β and θ. So the skill premium is irrelevant with γ.

A.4 Proof of Proposition 4

Proof. If Δ is very small, we can do a first-order expansion with respect to Δ around $\Delta = 0$.

$$\frac{1}{\theta(1-\beta)}[\Delta w^{\beta\theta} - \beta\theta \log w - \Delta w^{-\beta\theta}] + \frac{1}{1-\beta}[\log T_1^{\frac{1}{\theta}} L_1^{m\,\gamma} - \log T_2^{\frac{1}{\theta}} L_2^{m\,\gamma}] = \log w$$

$$\begin{aligned}
&\frac{1}{1-\alpha} \log w = -\frac{\beta}{1-\beta} \log w + \left(\frac{\gamma}{1-\beta} + 1\right) [\log L_1^m - \log L_2^m] \\
&+ \frac{1}{1-\beta}[\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] + \\
&\left(\frac{1}{\theta(1-\beta)}(w^{\beta\theta} - w^{-\beta\theta}) + \frac{\eta}{\eta+\zeta}(w^{-\beta\theta} - w^{\beta\theta} + \frac{1}{x} - x)\right) \Delta,
\end{aligned}$$

where

$$x = \left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta(1-\beta)}} w^{-\frac{\beta}{1-\beta}+\beta\theta} l^{m\,\frac{\gamma}{1-\beta}+1}.$$

That is,

$$\log w = \gamma \log l^m + \log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}} + \frac{1}{\theta}\Delta(w^{\beta\theta} - w^{-\beta\theta})$$

$$\begin{aligned}
&\left(\frac{1}{1-\alpha} + \frac{\beta}{1-\beta}\right) \log w = \left(\frac{\gamma}{1-\beta} + 1\right) \log l^m + \frac{1}{1-\beta}[\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] + \\
&\left(\frac{1}{\theta(1-\beta)}(w^{\beta\theta} - w^{-\beta\theta}) + \frac{\eta}{\eta+\zeta}(w^{-\beta\theta} - w^{\beta\theta} + \frac{1}{x} - x)\right) \Delta.
\end{aligned}$$

They lead to

$$(\gamma+1-\frac{\gamma}{1-\alpha})\log l^m = \frac{\alpha}{1-\alpha}[\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] + \left(\frac{1}{1-\alpha}+\frac{\beta}{1-\beta}\right)\frac{1}{\theta}(w^{\beta\theta}-w^{-\beta\theta})\Delta$$
$$-\left(\frac{1}{\theta(1-\beta)}(w^{\beta\theta}-w^{-\beta\theta}) + \frac{\eta}{\eta+\zeta}(w^{-\beta\theta}-w^{\beta\theta}+\frac{1}{x}-x)\right)\Delta.$$

That is,

$$(\gamma+1-\frac{\gamma}{1-\alpha})\log l^m = \frac{\alpha}{1-\alpha}[\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] + \frac{\alpha}{1-\alpha}\frac{1}{\theta}(w^{\beta\theta}-w^{-\beta\theta})\Delta$$
$$-\frac{\eta}{\eta+\zeta}(w^{-\beta\theta}-w^{\beta\theta}+\frac{1}{x}-x)\Delta.$$

By expanding w and x around $\Delta=0$, we finally arrive at

$$(\gamma+1-\frac{\gamma}{1-\alpha})\log l^m = \frac{\alpha}{1-\alpha}[\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] + [\frac{\alpha}{1-\alpha}\frac{1}{\theta}+\frac{\eta}{\eta+\zeta}](\hat{w}^{\beta\theta}-\hat{w}^{-\beta\theta})\Delta$$
$$+(\hat{x}-\frac{1}{\hat{x}})\Delta$$

where

$$\hat{x} = \left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta}(1-\beta)(1+\beta\theta+\frac{\alpha}{1-\alpha})} > 1,$$

and

$$\hat{w} = \left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta}\frac{1}{\gamma+1-\frac{\gamma}{1-\alpha}}} > 1$$

are solutions to x and w when $\Delta=0$.

Thus, managers will relocate to the bigger city with internet improvement. It is also easy to see that $\log w$ increases and $\log p$ increases locally with internet improvement. Moreover, we examine the role of agglomeration force γ from the following:

$$\frac{\partial \log l^m}{\partial \Delta} = \frac{1}{\gamma+1-\frac{\gamma}{1-\alpha}}\left[[\frac{\alpha}{1-\alpha}\frac{1}{\theta}+\frac{\eta}{\eta+\zeta}](\hat{w}^{\beta\theta}-\hat{w}^{-\beta\theta}) + (\hat{x}-\frac{1}{\hat{x}})\right].$$

A larger γ implies a bigger reallocation of high-skilled to the bigger city.

Also,

$$\frac{L_{p1}}{L_{p2}} = \frac{\eta w_1^{-1} T_1 (w_1^{\beta})^{-\theta} \Phi_1^{\frac{1}{\theta(1-\beta)}-1} L_1^{m\ \frac{\gamma}{1-\beta}+1} + \eta w_1^{-1} T_2 (w_1^{\beta})^{-\theta} \Delta \Phi_2^{\frac{1}{\theta(1-\beta)}-1} L_2^{m\ \frac{\gamma}{1-\beta}+1}}{\eta w_2^{-1} T_2 (w_2^{\beta})^{-\theta} \Phi_2^{\frac{1}{\theta(1-\beta)}-1} L_2^{m\ \frac{\gamma}{1-\beta}+1} + \eta w_2^{-1} T_1 (w_2^{\beta})^{-\theta} \Delta \Phi_1^{\frac{1}{\theta(1-\beta)}-1} L_1^{m\ \frac{\gamma}{1-\beta}+1}}.$$

Then

$$\frac{L_{p1}}{L_{p2}} = \frac{T_{11}/T_{22} w^{-1} (w^{\beta})^{-\theta} \Phi_{12}^{\frac{1}{\theta(1-\beta)}-1} (L_1^m / L_2^m)^{\frac{\gamma}{1-\beta}+1} + T_2^{-\theta}/T_2^{-\theta} w^{-1} (w^{\beta})^{-\theta} \Delta}{1 + T_1/T_2 \Delta \Phi_{12}^{\frac{1}{\theta(1-\beta)}-1} (L_1^m / L_2^m)^{\frac{\gamma}{1-\beta}+1}}$$

$$l_p = \frac{T_1/T_2 w^{-1} (w^{\beta})^{-\theta} \Phi_{12}^{\frac{1}{\theta(1-\beta)}-1} (L_1^m / L_2^m)^{\frac{\gamma}{1-\beta}+1} + T_2/T_2 w^{-1} (w^{\beta})^{-\theta} \Delta}{1 + T_1/T_2 \Delta \Phi_{12}^{\frac{1}{\theta(1-\beta)}-1} (L_1^m / L_2^m)^{\frac{\gamma}{1-\beta}+1}}$$

$$l_p = \frac{\left(\frac{T_1}{T_2}\right) w^{-1-\beta\theta} \left(\frac{\Phi_1}{\Phi_2}\right)^{\frac{1}{\theta(1-\beta)}-1} l^{m\ \frac{\gamma}{1-\beta}+1} + w^{-1-\beta\theta} \Delta}{1 + \left(\frac{T_1}{T_2}\right) \left(\frac{\Phi_1}{\Phi_2}\right)^{\frac{1}{\theta(1-\beta)}-1} l^{m\ \frac{\gamma}{1-\beta}+1} \Delta}.$$

When the cross-city communication cost is infinite,

$$\hat{l}_p = \left(\frac{T_1}{T_2}\right) \hat{w}^{-\frac{1}{1-\beta}} \left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta(1-\beta)}-1} \hat{l}_m^{\frac{\gamma}{1-\beta}+1} = \left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta} \frac{\frac{\alpha}{1-\alpha}}{\gamma+1-\frac{\gamma}{1-\alpha}}}.$$

That gives

$$\log l_p = \frac{\frac{\alpha}{1-\alpha}}{\gamma + 1 - \frac{\gamma}{1-\alpha}} [\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] - \left[\left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta(1-\beta)}} w^{-\frac{\beta}{1-\beta}+\beta\theta} l^{m\ \frac{\gamma}{1-\beta}+1} - \left(\frac{T_1}{T_2}\right)^{-\frac{1}{\theta(1-\beta)}} w^{\frac{\beta}{1-\beta}-\beta\theta} l^{m - \frac{\gamma}{1-\beta}-1} \right] \Delta$$

which leads to

$$\log l_p = \frac{\frac{\alpha}{1-\alpha}}{\gamma + 1 - \frac{\gamma}{1-\alpha}} [\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] - \left[\hat{w}^{1+\beta\theta} \hat{l}_p - \hat{w}^{-1-\beta\theta} \hat{l}_p^{-1}\right] \Delta.$$

It is clear that Δ increase will drive production workers from big cities to small cities. Moreover, we examine the role of agglomeration force γ from the following:

$$\frac{\partial \log l^p}{\partial \Delta} = -\left[\hat{w}^{1+\beta\theta} \hat{l}_p - \hat{w}^{-1-\beta\theta} \hat{l}_p^{-1}\right].$$

A larger γ implies a bigger reallocation of the low skilled to the smaller city. ■

B Skill Premium and Welfare in the Two-city Case

This section presents welfare analysis for a two-city version of the baseline model presented in Section 3.4.

We first do a first-order expansion around $\Delta = 0$. The skill premium is

$$\begin{aligned} sp &= \frac{\gamma}{1-\beta}\log L_1^m + \frac{1}{(1-\beta)\theta}\log\Phi_1 - \log w_1 \\ &= \frac{\gamma}{1-\beta}\log L_1^m + \frac{1}{(1-\beta)\theta}[\log T_1 + \log(1+\Delta w^{\beta\theta})] - \frac{1}{1-\beta}\log w_1 \\ &= \frac{\gamma}{1-\beta}\log L_1^m + \frac{1}{(1-\beta)\theta}[\log T_1 + (1+\Delta \hat{w}^{\beta\theta})] - \frac{1}{1-\beta}\log w_1, \end{aligned} \tag{43}$$

where $w = w_1/w_2$.

Note that the market-clearing condition for production workers is

$$T_1\Phi_1^{\frac{1}{\theta(1-\beta)}-1}L_1^{m\,\frac{\gamma}{1-\beta}+1}(w_1^{-\beta\theta-1}+\Delta w_2^{-\beta\theta-1}) + T_2\Phi_2^{\frac{1}{\theta(1-\beta)}-1}L_2^{m\,\frac{\gamma}{1-\beta}+1}(w_2^{-\beta\theta-1}+\Delta w_1^{-\beta\theta-1}) = \frac{L_p}{\eta}.$$

That is,

$$\begin{aligned} & T_1^{\frac{1}{\theta(1-\beta)}}(w_1^{-\beta\theta}+\Delta w_2^{-\beta\theta})^{\frac{1}{\theta(1-\beta)}-1}L_1^{m\,\frac{\gamma}{1-\beta}+1}(w_1^{-\beta\theta-1}+\Delta w_2^{-\beta\theta-1}) \\ & + T_2^{\frac{1}{\theta(1-\beta)}}(\Delta w_1^{-\beta\theta}+w_2^{-\beta\theta})^{\frac{1}{\theta(1-\beta)}-1}L_2^{m\,\frac{\gamma}{1-\beta}+1}(\Delta w_1^{-\beta\theta-1}+w_2^{-\beta\theta-1}) = \frac{L_p}{\eta}. \end{aligned} \tag{44}$$

Rearrange to get

$$\begin{aligned} & T_1^{\frac{1}{\theta(1-\beta)}}(1+\Delta w^{\beta\theta})^{\frac{1}{\theta(1-\beta)}-1}(1+\Delta w^{\beta\theta+1}) \\ & + T_2^{\frac{1}{\theta(1-\beta)}}(\Delta + w^{\beta\theta})^{\frac{1}{\theta(1-\beta)}-1}(L_1^m/L_2^m)^{-\frac{\gamma}{1-\beta}-1}(\Delta + w^{\beta\theta+1}) = \frac{L_p}{\eta}w_1^{\frac{1}{1-\beta}}L_1^{m-\frac{\gamma}{1-\beta}-1}. \end{aligned} \tag{45}$$

Recall that

$$(\gamma+1-\frac{\gamma}{1-\alpha})\log l^m = \frac{\alpha}{1-\alpha}[\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] + [\frac{\alpha}{1-\alpha}\frac{1}{\theta} + \frac{\eta}{\eta+\zeta}](\hat{w}^{\beta\theta} - \hat{w}^{-\beta\theta})\Delta$$
$$+ (\hat{x} - \frac{1}{\hat{x}})\Delta$$

where

$$\hat{x} = \left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta}(1-\beta)(1+\beta\theta+\frac{\alpha}{1-\alpha})} > 1,$$

and

$$\hat{w} = \left(\frac{T_1}{T_2}\right)^{\frac{1}{\theta}\frac{1}{\gamma+1-\frac{\gamma}{1-\alpha}}} > 1$$

are solutions to x and w when $\Delta = 0$. Therefore,

$$\begin{aligned}(-\frac{\gamma}{1-\beta}-1)\log l^m \quad &= -\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}\frac{\alpha}{1-\alpha}[\log T_1^{\frac{1}{\theta}} - \log T_2^{\frac{1}{\theta}}] - \frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}[\frac{\alpha}{1-\alpha}\frac{1}{\theta}+\frac{\eta}{\eta+\zeta}] \\ &\quad (\hat{w}^{\beta\theta}-\hat{w}^{-\beta\theta})\Delta - \frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}(\hat{x}-\frac{1}{\hat{x}})\Delta.\end{aligned}$$

This implies that

$$l_m^{-\frac{\gamma}{1-\beta}-1} = (\frac{T_1}{T_2})^{-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}\frac{\alpha}{1-\alpha}\frac{1}{\theta}}\left[1-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}[\frac{\alpha}{1-\alpha}\frac{1}{\theta}+\frac{\eta}{\eta+\zeta}](\hat{w}^{\beta\theta}-\hat{w}^{-\beta\theta})\Delta - \frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}(\hat{x}-\frac{1}{\hat{x}})\Delta\right].$$

As a result,

$$\begin{aligned}&T_1^{\frac{1}{\theta(1-\beta)}}[1+(\hat{w}^{\beta\theta}(\frac{1}{\theta(1-\beta)}-1)+\hat{w}^{\beta\theta+1})\Delta] + T_2^{\frac{1}{\theta(1-\beta)}}\hat{w}^{\beta\theta(\frac{1}{\theta(1-\beta)}-1)}\hat{w}^{\beta\theta+1}(\frac{T_1}{T_2})^{-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}\frac{\alpha}{1-\alpha}\frac{1}{\theta}}\times \\ &(1+\hat{w}^{-\beta\theta}\Delta+\hat{w}^{-\beta\theta-1}\Delta - \frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}[\frac{\alpha}{1-\alpha}\frac{1}{\theta}+\frac{\eta}{\eta+\zeta}](\hat{w}^{\beta\theta}-\hat{w}^{-\beta\theta})\Delta - \frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}(\hat{x}-\frac{1}{\hat{x}})\Delta) \\ &= \frac{L_p}{\eta}w_1^{\frac{1}{1-\beta}}L_1^{m-\frac{\gamma}{1-\beta}-1}. \qquad (46)\end{aligned}$$

We can then obtain that

$$\frac{1}{1-\beta}\log w_1 - (\frac{\gamma}{1-\beta}+1)\log L_1^m = constant + \frac{A}{T_1^{\frac{1}{\theta(1-\beta)}} + T_2^{\frac{1}{\theta(1-\beta)}}w^{\frac{1}{1-\beta}}(\frac{T_1}{T_2})^{-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}\frac{\alpha}{1-\alpha}\frac{1}{\theta}}}\Delta,$$

where

$$A = T_1^{\frac{1}{\theta(1-\beta)}}(\hat{w}^{\beta\theta}(\frac{1}{\theta(1-\beta)}-1)+\hat{w}^{\beta\theta+1})+T_2^{\frac{1}{\theta(1-\beta)}}\hat{w}^{\beta\theta(\frac{1}{\theta(1-\beta)}-1)}\hat{w}^{\beta\theta+1}(\frac{T_1}{T_2})^{-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}\frac{\alpha}{1-\alpha}\frac{1}{\theta}}\times$$
$$\left[\hat{w}^{-\beta\theta}+\hat{w}^{-\beta\theta-1}-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}[\frac{\alpha}{1-\alpha}\frac{1}{\theta}+\frac{\eta}{\eta+\zeta}](\hat{w}^{\beta\theta}-\hat{w}^{-\beta\theta})-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}(\hat{x}-\frac{1}{\hat{x}})\right]. \tag{47}$$

Since the skill premium is

$$\frac{1}{(1-\beta)\theta}[\log T_1+(1+\Delta\hat{w}^{\beta\theta})]-(\frac{1}{1-\beta}\log w_1-\frac{\gamma}{1-\beta}\log L_1^m),$$

we know that the sign of the skill premium is the same as the sign of

$$T_1^{\frac{1}{\theta(1-\beta)}}(\hat{w}^{\beta\theta}-\hat{w}^{\beta\theta+1})+T_2^{\frac{1}{\theta(1-\beta)}}\hat{w}^{\beta\theta(\frac{1}{\theta(1-\beta)}-1)}\hat{w}^{\beta\theta+1}(\frac{T_1}{T_2})^{-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}\frac{\alpha}{1-\alpha}\frac{1}{\theta}}\times\frac{1}{(1-\beta)\theta}\hat{w}^{\beta\theta}-$$
$$T_2^{\frac{1}{\theta(1-\beta)}}\hat{w}^{\beta\theta(\frac{1}{\theta(1-\beta)}-1)}\hat{w}^{\beta\theta+1}(\frac{T_1}{T_2})^{-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}\frac{\alpha}{1-\alpha}\frac{1}{\theta}}\times$$
$$\left[\hat{w}^{-\beta\theta}+\hat{w}^{-\beta\theta-1}-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}[\frac{\alpha}{1-\alpha}\frac{1}{\theta}+\frac{\eta}{\eta+\zeta}](\hat{w}^{\beta\theta}-\hat{w}^{-\beta\theta})-\frac{\frac{\gamma}{1-\beta}+1}{\gamma+1-\frac{\gamma}{1-\alpha}}(\hat{x}-\frac{1}{\hat{x}})\right]. \tag{48}$$

The sign can be positive or negative.

Managers' welfare is

$$\frac{\gamma}{1-\beta}\log L_1^m+\frac{1}{\theta(1-\beta)}\log\Phi_1-(1-\alpha)\log p_1. \tag{49}$$

Since housing price is given by

$$\log p_1=\log(1-\alpha)+\log W_1-\log H_1, \tag{50}$$

where

$$W_1=w_1L_1^p+\zeta L_1^{m\,\frac{\gamma}{1-\beta}}\Phi_1^{\frac{1}{\theta(1-\beta)}}L_1^m, \tag{51}$$

managers' welfare is rewritten as

$$\frac{\gamma}{1-\beta}\log L_1^m + \frac{1}{\theta(1-\beta)}\log\Phi_1 - (1-\alpha)\log w_1 - \log(L_p^1 + \zeta L_1^{m\,\frac{\gamma}{1-\beta}}\Phi_1^{\frac{1}{\theta(1-\beta)}} L_1^m/w_1)$$
$$= skill\_premium + \alpha\log w_1 - \log(L_p^1 + \zeta L_1^{m\,\frac{\gamma}{1-\beta}}\Phi_1^{\frac{1}{\theta(1-\beta)}} L_1^m/w_1). \tag{52}$$

Similarly, we can substitute expressions for the skill premium, $\log w_1$, L_1^m, and L_1^p derived above, to get $\frac{\partial \text{ managers' welfare}}{\partial\Delta}$ evaluated at $\Delta = 0$. Finally, production workers' welfare is equal to managers' welfare minus the skill premium.

C Equilibrium Characterization

In this section, we show, using Banach fixed point theorem, a set of sufficient conditions under which the equilibrium exists and is unique. For simplicity of exposition, we denote $\Delta_{nc} = \tau_{nc}^{-\theta}$. For the derivation of analytic results, we follow the conventional literature and adopt the following parametric assumption for the city-level agglomeration forces for managers (see, e.g., Allen and Arkolakis, 2014):

$$f(L) = L^\gamma, \quad \text{where } \gamma > 0.$$

Combining this assumption and the equilibrium housing prices in (11), we can rewrite the indifference conditions for workers and managers in (15) and (16) as

$$\frac{\gamma}{1-\beta}\log\frac{L_n^m}{L_c^m} + \frac{1}{\theta(1-\beta)}\log\frac{\Phi_n}{\Phi_c} = (1-\alpha)\log\frac{p_n}{p_c} = \log\frac{w_n}{w_c} \tag{53}$$

and

$$\left(\frac{w_n}{w_c}\right)^{\frac{1}{1-\alpha}} = \frac{\zeta(L_n^{m\,\gamma\theta}\Phi_n)^{\frac{1}{\theta(1-\beta)}} L_n^m + \sum_k \eta\left(T_k(\tau_{kn}w_n^\beta)^{-\theta}\right)\Phi_k^{\frac{1}{\theta(1-\beta)}-1}[L_k^m]^{\frac{\gamma}{1-\beta}+1}}{\zeta(L_c^{m\,\gamma\theta}\Phi_c)^{\frac{1}{\theta(1-\beta)}} L_c^m + \sum_k \eta\left(T_k(\tau_{kc}w_c^\beta)^{-\theta}\right)\Phi_k^{\frac{1}{\theta(1-\beta)}-1}[L_k^m]^{\frac{\gamma}{1-\beta}+1}}\frac{H_c}{H_n}. \tag{54}$$

We can then solve for $\frac{w_n}{w_c}$ and $\frac{L_n^m}{L_c^m}$ from the above two equations. In a special case with no agglomeration force $\gamma = 0$, we can first solve worker's wage w_n from equation (53) and then the number of managers L_n^m from equation (54).

The relative number of production workers in city n and city c is given by

$$\frac{L_n^p}{L_c^p} = \frac{\sum_k \eta w_n^{-1} \left(T_k(\tau_{kn} w_n^{\beta})^{-\theta}\right) \Phi_k^{\frac{1}{\theta(1-\beta)}-1} [L_k^m]^{\frac{\gamma}{1-\beta}+1}}{\sum_k \eta w_c^{-1} \left(T_k(\tau_{kc} w_c^{\beta})^{-\theta}\right) \Phi_k^{\frac{1}{\theta(1-\beta)}-1} [L_k^m]^{\frac{\gamma}{1-\beta}+1}}. \tag{55}$$

Finally, the skill premium, defined as the log difference between the manager's and production worker's expected income, is given by:

$$\log E[\pi_n] - \log w_n = \frac{\gamma}{1-\beta} \log L_n^m + \frac{1}{(1-\beta)\theta} \log \Phi_n - \log w_n. \tag{56}$$

Proposition 5 *(Existence and Uniqueness) If there exists an $aux \in \mathbb{R}$ such that*

$$\rho = |\frac{-\beta\theta + aux}{1/(1-\alpha) + aux}| + |\frac{\frac{1-\beta}{\gamma} + 1}{1/(1-\alpha) + aux}| + |(1 + \frac{1}{\theta\gamma})(\frac{-\beta\theta}{1/(1-\alpha) + aux})| < 1,$$

then the spatial equilibrium exists and is unique.

Proof. We obtain the following equation from the definition of Φ_n in (8):

$$\begin{bmatrix} \Delta_{11}T_1 & \Delta_{12}T_1 & \Delta_{13}T_1 & \dots & \Delta_{1N}T_1 \\ \Delta_{21}T_2 & \Delta_{22}T_2 & \Delta_{23}T_2 & \dots & \Delta_{2N}T_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Delta_{N1}T_N & \Delta_{N2}T_N & \Delta_{N3}T_2 & \dots & \Delta_{NN}T_N \end{bmatrix} \begin{bmatrix} w_1^{-\beta\theta} \\ w_2^{-\beta\theta} \\ \vdots \\ w_N^{-\beta\theta} \end{bmatrix} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_N \end{bmatrix}. \tag{57}$$

From equation (53), we get

$$L_n^{m\,\frac{\gamma}{1-\beta}} \propto w_n \Phi_n^{-\frac{1}{\theta(1-\beta)}}. \tag{58}$$

Therefore, up to a constant, and using results from equation (54), we can rewrite the matrix as

$$
\begin{bmatrix}
\zeta\Phi_1^{\frac{1}{\theta(1-\beta)}} + \eta T_1\Delta_{11}w_1^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1} & \eta T_2\Delta_{21}w_1^{-\beta\theta}\Phi_2^{\frac{1}{\theta(1-\beta)}-1} & \dots & \eta T_N\Delta_{N1}w_1^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1} \\
\eta T_1\Delta_{12}w_2^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1} & \zeta\Phi_2^{\frac{1}{\theta(1-\beta)}} + \eta T_2\Delta_{22}w_2^{-\beta\theta}\Phi_2^{\frac{1}{\theta(1-\beta)}-1} & \dots & \eta T_N\Delta_{N2}w_2^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1} \\
\vdots & \vdots & \ddots & \vdots \\
\eta T_1\Delta_{1N}w_N^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1} & \eta T_2\Delta_{2N}w_N^{-\beta\theta}\Phi_2^{\frac{1}{\theta(1-\beta)}-1} & \dots & \zeta\Phi_N^{\frac{1}{\theta(1-\beta)}} + \eta T_N\Delta_{NN}w_N^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1}
\end{bmatrix}
$$

$$
\times \begin{bmatrix}
w_1^{\frac{1-\beta}{\gamma}+1}\Phi_1^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\
w_2^{\frac{1-\beta}{\gamma}+1}\Phi_2^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\
\vdots \\
w_N^{\frac{1-\beta}{\gamma}+1}\Phi_N^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}}
\end{bmatrix}
=
\begin{bmatrix}
w_1^{\frac{1}{1-\alpha}}H_1 \\
w_2^{\frac{1}{1-\alpha}}H_2 \\
\vdots \\
w_N^{\frac{1}{1-\alpha}}H_N
\end{bmatrix}. \tag{59}
$$

Multiply both sides by w_n^{aux}, where $aux \in \mathcal{R}$ is an auxiliary parameter, and we get

$$
\begin{bmatrix}
\zeta\Phi_1^{\frac{1}{\theta(1-\beta)}}w_1^{aux} + \eta T_1\Delta_{11}w_1^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1}w_1^{aux} & \dots & \eta T_N\Delta_{N1}w_1^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1}w_1^{aux} \\
\eta T_1\Delta_{12}w_2^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1}w_2^{aux} & \dots & \eta T_N\Delta_{N2}w_2^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1}w_2^{aux} \\
\vdots & \ddots & \vdots \\
\eta T_1\Delta_{1N}w_N^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1}w_N^{aux} & \dots & \zeta\Phi_N^{\frac{1}{\theta(1-\beta)}}w_N^{aux} + \eta T_N\Delta_{NN}w_N^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1}w_N^{aux}
\end{bmatrix}
$$

$$
\times \begin{bmatrix}
w_1^{\frac{1-\beta}{\gamma}+1}\Phi_1^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\
w_2^{\frac{1-\beta}{\gamma}+1}\Phi_2^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\
\vdots \\
w_N^{\frac{1-\beta}{\gamma}+1}\Phi_N^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}}
\end{bmatrix}
=
\begin{bmatrix}
w_1^{\frac{1}{1-\alpha}+aux}H_1 \\
w_2^{\frac{1}{1-\alpha}+aux}H_2 \\
\vdots \\
w_N^{\frac{1}{1-\alpha}+aux}H_N
\end{bmatrix}. \tag{60}
$$

Denote $x_n = \left(\frac{1}{1-\alpha} + aux\right)\log w_n$ and $\boldsymbol{x} = (x_1, x_2, ...x_N)'$. Then

$$
F_i(\boldsymbol{x}) = \log[\sum_j \frac{1}{H_i}\eta\exp(x_i)^{\frac{-\beta\theta+aux}{1/(1-\alpha)+aux}}T_j\Delta_{ji}\exp(x_j)^{\frac{\frac{1-\beta}{\gamma}+1}{1/(1-\alpha)+aux}}\left(\Phi_j^a\right)^{-1-\frac{1}{\theta\gamma}}
$$
$$
+ \zeta\frac{1}{H_i}\exp(x_i)^{\frac{\frac{1-\beta}{\gamma}+1+aux}{1/(1-\alpha)+aux}}\left(\Phi_i^a\right)^{-\frac{1}{\theta\gamma}}],
$$

where

$$
\Phi_j^a = T_j\sum_k \Delta_{jk}\exp(x_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}.
$$

Finally, we can show that

$$d(F(\boldsymbol{x}), F(\boldsymbol{y})) = \max_i |F_i(x) - F_i(y)|$$

$$= \max_i \log[\sum_j \lambda_j exp(x_i - y_i)^{\frac{-\beta\theta + aux}{1/(1-\alpha)+aux}} exp(x_j - y_j)^{\frac{\frac{1-\beta}{\gamma}+1}{1/(1-\alpha)+aux}}$$

$$\times \left(\frac{T_j \sum_k \Delta_{jk} \exp(x_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}}{T_j \sum_k \Delta_{jk} \exp(y_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}} \right)^{-1-\frac{1}{\theta\gamma}} + \lambda_{N+1} exp(x_i - y_i)^{\frac{\frac{1-\beta}{\gamma}+1+aux}{1/(1-\alpha)+aux}} \times \left(\frac{T_j \sum_k \Delta_{jk} \exp(x_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}}{T_j \sum_k \Delta_{jk} \exp(y_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}} \right)^{-\frac{1}{\theta\gamma}}$$

where $\sum_j \lambda_j + \lambda_{N+1} = 1$, $\lambda_j \geq 0$, $\lambda_{N+1} \geq 0$. Note that

$$\frac{T_j \sum_k \Delta_{jk} \exp(x_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}}{T_j \sum_k \Delta_{jk} \exp(y_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}} \leq \sum_k \omega_k exp(x_k - y_k)^{\frac{-\beta\theta}{1/(1-\alpha)+aux}}$$

where

$$\sum_k \omega_k = 1, \omega_k \geq 0.$$

Therefore, we get that

$$d(F(\boldsymbol{x}), F(\boldsymbol{y})) \leq \rho \cdot max_k |x_k - y_k| = \rho \cdot d(\boldsymbol{x}, \boldsymbol{y}),$$

where

$$\rho = |\frac{-\beta\theta + aux}{1/(1-\alpha) + aux}| + |\frac{\frac{1-\beta}{\gamma} + 1}{1/(1-\alpha) + aux}| + |(1 + \frac{1}{\theta\gamma})(\frac{-\beta\theta}{1/(1-\alpha) + aux})|.$$

If there exists a real number aux such that $\rho < 1$, using Banach fixed-point theorem, the equilibrium exists and is unique. ■

D Demand for Production Workers

We derive the labor demand for production workers given by equation (19).

Denote $X_{nc} = \frac{\bar{a}_{nc}}{\tau_{nc} w_c^\beta}$. Then from Proposition 1, we get

$$G_{nc}(x) = Pr(X_{nc} \leq x) = Pr(A_{nc} \leq \tau_{nc} w_c^\beta x) = e^{-T_n (\tau_{nc} w_c^\beta)^{-\theta} x^{-\theta}}.$$

The joint distribution whereby a manager from city n locates her production team in city c and that $\frac{\bar{a}_{nc}}{\tau_{nc} w_c^{\beta}} = x$ is

$$Pr(argmax_k \frac{\bar{a}_{nk}}{\tau_{nk} w_k^{\beta}} = c \cap \frac{\bar{a}_{nc}}{\tau_{nc} w_c^{\beta}} = x) = \theta T_n (\tau_{nc} w_c^{\beta})^{-\theta} x^{-\theta-1} e^{-\Phi_n x^{-\theta}}.$$

Given $l_{nc} = \beta^{\frac{1}{1-\beta}} w_c^{-1} [f(L_n^m)]^{\frac{1}{1-\beta}} \left[\frac{\bar{a}_{nc}}{\tau_{nc} w_c^{\beta}}\right]^{\frac{1}{1-\beta}}$, we have

$$\begin{aligned} L_{nc}^{p} &= \beta^{\frac{1}{1-\beta}} \left(T_n (\tau_{nc} w_c^{\beta})^{-\theta}\right) w_c^{-1} [f(L_n^m)]^{\frac{1}{1-\beta}} L_n^m \left[\int_0^{\infty} \left(\theta x^{-\theta-1} e^{-\Phi_n x^{-\theta}}\right) x^{\frac{1}{1-\beta}} dx\right], \\ &= \eta w_c^{-1} \left(T_n (\tau_{nc} w_c^{\beta})^{-\theta}\right) \Phi_n^{\frac{1}{\theta(1-\beta)}-1} \left[[f(L_n^m)]^{\frac{1}{1-\beta}} L_n^m\right] \end{aligned}$$

where $\eta = \beta^{\frac{1}{1-\beta}} \int_0^{\infty} y^{-\frac{1}{\theta(1-\beta)}} e^{-y} dy$.

E Heterogeneity Analysis across Industries

We provide further evidence in this section to link the observed increase in spatial segregation with our proposed mechanism of increasing production fragmentation across U.S. cities. In particular, we show that this pattern of segregation across space at the industry level is closely related to production fragmentation activities in the U.S. economy. Fort (2017) documents that firms' adoption of communications technology facilitates their sourcing, particularly from domestic suppliers. If the observed segregation is linked to greater sourcing of tasks, one would expect that industries experiencing more sourcing would also undergo greater skill segregation. Figure 9 confirms this hypothesis by illustrating the relationship between the change in the KM index and the fraction of plants that engage in sourcing activities. These are measured by the purchases of contract manufacturing services (CMS) from other plants (within its own company or from another company) in each of the 86 four-digit NAICS manufacturing industries.[26] For example, the computers and related equipment industry, which has a very high sourcing index (50% of plants source from another plant), features a relatively large increase in the KM index; in contrast, the bakery product industry has a very

[26] Fort (2017) provides this measure at the four-digit NAICS level. We employ the industry code crosswalk between census industries and NAICS industries provided by the Census Bureau, so that each census industry is assigned to the corresponding NAICS code. If one Census industry corresponds to multiple NAICS codes, we calculate the simple average of the fragmentation indices among those NAICS codes as that census industry's fragmentation index. We are left with 67 census industries. See Table 13 in the Appendix for a detailed list of industries and their change in KM index and Fort (2017) index.

low sourcing index (8% of plants source from another plant) exhibits a slight decrease in the KM index.

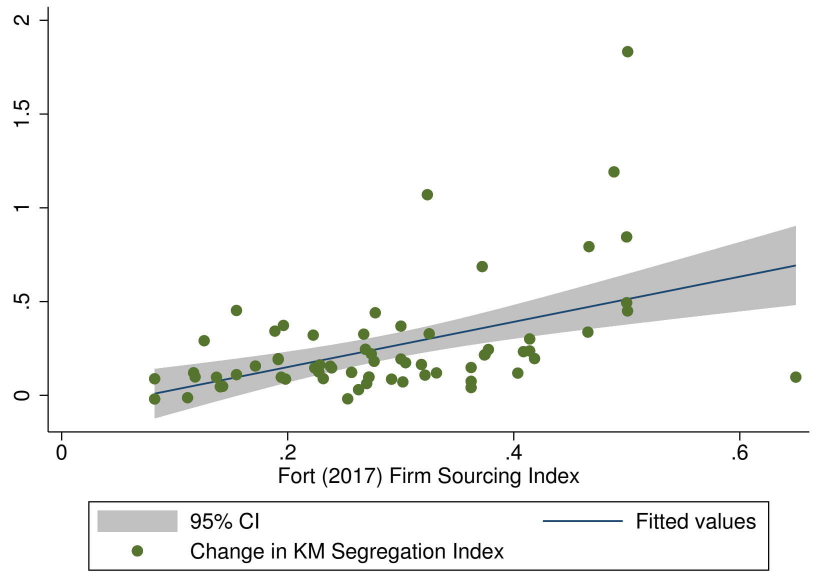


Figure 9: Change in KM Skill Segregation Index and Fragmentation Index Across Sectors

Notes: Each point denotes an industry. The shaded area displays the 95% confidence band around the point estimates for the slope. The correlation between change in KM skill segregation index and Fort (2017) sourcing index is 0.47.

In summary, we show that ihe extent of segregation varies across industries systematically, matching the cross-industry heterogeneity in the extent of production fragmentation.

F Internet Quality and Bilateral Goods Trade

We draw data from the Commodity Flow Survey (CFS) 2012, which is publicly accessible. The survey records the value of shipments from region i to region j, where each region is defined as a combined statistical area (CSA) in the U.S.. Supplementing these data with the geographic information and internet quality data, we run the following regression:

$$\log(shipment_{ij}) = \beta_0 + \beta_1 log(distance_{ij}) + \beta_2 q_i * q_j + X_{ij} + \epsilon_{ij}, \tag{61}$$

where the dependent variable is the logarithm of the total value of the shipment from CSA i to CSA j, $distance_{ij}$ is the great circle distance, and q_i is internet quality at the CSA i. We perform the regression with and without location fixed effects.

Table 9 reports the results. In Column (1), we don't include origin and destination fixed effects, but include internet quality in i and j as additional explanatory variables. We find that while bilateral distance significantly reduces trade, the impact of the internet on bilateral goods trade is not statistically significant. In Column (2), we include both origin and destination fixed effects. Similarly, the point estimate for bilateral internet connection term $q_i * q_j$ is even smaller and is still statistically insignificant.

Dependent variable	log(*shipment*)	log(*shipment*)
	(1)	(2)
log (distance)	-1.236***	-1.239***
	(.0026)	(.027)
$q_i * q_j$	.058	.039
	(.094)	(.053)
q_i	.489	
	(.349)	
q_j	.379	
	(.356)	
CSA Fixed Effects	No	Yes
N	4,801	4,801

Table 9: Gravity Equation Estimates for Trade in Goods

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

G Solution Algorithm for the Spatial Equilibrium

For a system of $N \geq 1$ cities, given a set of bilateral fragmentation cost τ_{nc} (or $\Delta_{nc} = \tau_{nc}^{-\theta}$), technology T_n, housing supply H_n, and aggregate labor supply L^p and L^m, we use the following iteration methods (global methods) to solve the spatial equilibrium:

(1) Pick up an initial set of production worker's (relative) wage $wr^{(0)}$ (a vector of $N-1$ dimensions), where $wr^{(j)} = w^{(j)}(1:N-1)/w^{(j)}(N)$. For example, all $N-1$ elements of $wr^{(0)}$ are set to 1.

(2) Pick up a real number aux. If the code doesn't converge, pick up another aux and retry the steps that follow. Start from $j = 0$.

(3) From equation (8), which defines Φ_n, we calculate relative Φr (a vector of $N-1$ dimensions), where $\Phi r^{(j)} = \Phi^{(j)}(1:N-1)/\Phi^{(j)}(N)$.

(4) Substitute vectors $wr^{(j)}$, $\Phi r^{(j)}$ into the left-hand side of equation (60) to update vector w on the right-hand side. The updated relative wage is written as $wr^{(j+1)}$.

(5) Check convergence, if $\max_{k=1,2,\ldots,N-1} |wr^{(j)}(k) - wr^{(j+1)}(k)| \leq \epsilon$, where, for example, $\epsilon = 10^{-6}$. Then stop. Otherwise, go back to step (3) and continue.

(6) From equation (53), which gives managers' spatial distribution, we get the relative number of managers Lr^m (a vector of $N-1$ dimensions), where $Lr^m = L^m(1:N-1)/L^m(N)$.

(7) With the total supply of managers L^m and relative number of managers Lr^m, we obtain the equilibrium spatial distribution of managers. Equation (55) gives the relative number of production workers. With the total supply of production workers L^p, we get the equilibrium spatial distribution of production workers.

(8) The level of worker wage w is solved using the market-clearing condition for production workers; see equations (18) and (19).

(9) With the expected income of managers living in city n (see equation (13)), worker wage, number of managers, and number of production workers, we can add up the total income W_n in a city. Combined with information on housing supply H_n, we get housing price $p_n = (1-\alpha)W_n/H_n$.

(10) Worker utility and managers' expected utility are derived as $w_n/p_n^{1-\alpha}$ and $E[\pi_n]/p_n^{1-\alpha}$.

H Sensitivity Analysis γ

Recall that γ in our model is calibrated to match the (average) worker wage elasticity with respect to city size. In our equilibrium analysis in Proposition 4, we argue that larger γ will imply more skill redistribution to larger cities (cities with greater technology parameter T_n), given the same magnitude of communication cost reduction. We next evaluate how our key quantitative results are sensitive to other values of γ.

In the baseline analysis, we set $\gamma = 0.010$. We change its value to $\gamma = 0.005$ and 0.015 to conduct the sensitivity analysis.[27] We report the quantitative results in Table 10, including the slope of regressing the change in the high-skill share on city size and welfare (real consumption) change for both managers and production workers.

It turns out that changing γ has very little impact on the results. The slope of regressing the change in the high-skill share on city size actually becomes slightly smaller when γ increases, if anything. This is precisely because of the re-calibration of T_n.[28] Equation (36) tells us that

$$\frac{1}{\theta(1-\beta)} \log \frac{T_n}{T_c} = \log \frac{w_n}{w_c} - \frac{1}{\theta(1-\beta)} \log \frac{\sum_k (\tau_{nk} w_k^{\beta})^{-\theta}}{\sum_k (\tau_{ck} w_c^{\beta})^{-\theta}} - \frac{\gamma}{1-\beta} \log \frac{L_n^m}{L_c^m}. \tag{62}$$

Without loss of generality, consider a big city n, which usually has a relatively larger worker wage w_n and more supply of high-skilled L_n^m. A larger γ will reduce the value of the calibrated T_n (T_n/T_c). This means that big cities' comparative advantage is weakened. Therefore, the impact of a reduction in communication cost will be smaller. The two forces – a larger γ and a smaller T_n for big cities – will offset each other to some extent. As a result, changing γ has very little impact on how communication cost reduction affects the skill redistribution. Similarly, the welfare increases (change in log real consumption) for both managers and production workers change little with the value of γ.

γ	slope of skilled share change w.r.t. city size	Δ workers' welfare	Δ managers' welfare
0.005	0.0030	3.62%	3.39%
0.010 (benchmark)	0.0031	3.62%	3.39%
0.015	0.0032	3.62%	3.39%

Table 10: Sensitivity of Changing the Strength of Agglomeration γ

Notes: This table shows how different values of the strength of agglomeration γ impacts our quantitative results. Δ workers' welfare and Δ managers' welfare are changes in log real consumption.

[27] When we change the value of γ, we also need to redo the calibration of technology parameter T_n derived from equation (36), since it will be affected by the value of γ.

[28] Table 11 in the Appendix reports quantitative results without recalibrating T_n as a thought experiment.

I Sensitivity Analysis δ_I

In our baseline exercise, we directly take the estimate of δ_I, the elasticity of fragmentation cost with respect to bilateral internet quality, from a gravity equation test. We emphasize the role of internet in shaping bilateral fragmentation cost by varying the value of δ_I in this section.

The baseline exercise gives the value $\delta_I = -0.010$. We try other two values, -0.005 and -0.015, which are half and double the baseline value. The results on skill relocation and welfare are summarized in Table 11.

Changing how the internet shapes fragmentation cost δ_I drastically reshapes skill relocation and welfare change. When $\delta_I = -0.005$, the change in the elasticity of the high-skilled share with respect to city size drops to 0.0018 from 0.0030 in the baseline, which is a 40% drop. Similar magnitudes of drops occur in both production workers' and managers' real consumption levels. When $\delta_I = -0.015$, the opposite happens. The elasticity of the high-skill share with respect to city size increases to 0.0037 from 0.0030, which is around a 23% increase. The welfare changes are even larger for both production workers and managers are of the similar scales of increase.

δ_I	slope of skilled share change w.r.t. city size	Δ workers' welfare	Δ managers' welfare
-0.005	0.0018	2.03%	1.89%
-0.010 (benchmark)	0.0031	3.62%	3.39%
-0.015	0.0037	4.60%	4.31%

Table 11: Sensitivity of Changing the Elasticity of Fragmentation Cost w.r.t. Internet Quality δ_I

Notes: This table shows how different values of the elasticity of fragmentation cost with respect to internet quality δ_I impacts our quantitative results. Δ workers' welfare and Δ managers' welfare are changes in log real consumption.

J Derivation of Production Function S_{nc}

We have the following conditional probability derived in Appendix D

$$Pr(\frac{\bar{a}_{nc}}{\tau_{nc}w_c^{\beta}} = x|argmax_k \frac{\bar{a}_{nk}}{\tau_{nc}w_c^{\beta}} = c) = \frac{Pr(\frac{\bar{a}_{nc}}{\tau_{nc}w_c^{\beta}} = x, argmax_k \frac{\bar{a}_{nk}}{\tau_{nc}w_c^{\beta}} = c)}{Pr(argmax_k \frac{\bar{a}_{nk}}{\tau_{nc}w_c^{\beta}} = c)} = \theta x^{-\theta-1} e^{-\Phi_n x^{-\theta}} \Phi_n. \tag{63}$$

Then

$$\begin{aligned} Pr(\bar{a}_{nc} = \bar{a}|argmax_k \frac{\bar{a}_{nk}}{\tau_{nc}w_c^{\beta}} = c) =& \theta \left(\frac{\bar{a}}{\tau_{nc}w_c^{\beta}}\right)^{-\theta-1} e^{-\Phi_n \left(\frac{\bar{a}}{\tau_{nc}w_c^{\beta}}\right)^{-\theta}} \Phi_n \frac{1}{\tau_{nc}w_c^{\beta}} \\ =& \theta \bar{a}^{-\theta-1} e^{-\Phi_n \tau_{nc}^{\theta} w_c^{\beta\theta} \bar{a}^{-\theta}} \Phi_n \tau_{nc}^{\theta} w_c^{\beta\theta}. \end{aligned} \tag{64}$$

Therefore,

$$Pr(\bar{a}_{nc} \le \bar{a}|argmax_k \frac{\bar{a}_{nk}}{\tau_{nc}w_c^{\beta}} = c) = e^{-\Phi_n \tau_{nc}^{\theta} w_c^{\beta\theta} \bar{a}^{-\theta}}. \tag{65}$$

Denote $\iota(\bar{a}) = Pr(\bar{a}_{nc} = \bar{a}|argmax_k \frac{\bar{a}_{nk}}{\tau_{nc}w_c^{\beta}} = c)$.

Suppose that there are L_{nc}^m managers and L_{nc}^p workers that form production teams between city n and city c, with n the origin city. For a team with managers productivity $\bar{a}$, her objective function is

$$\max_l f(L_n^m) \frac{\bar{a}}{\tau_{nc}} l^{\beta} - w_c l.$$

We get that

$$l(\bar{a}) = \left(\frac{\beta f(L_n^m)\bar{a}}{\tau_{nc}w_c}\right)^{\frac{1}{1-\beta}} = \kappa_{nc} \bar{a}^{\frac{1}{1-\beta}},$$

where $\kappa_{nc} = \beta^{\frac{1}{1-\beta}} (L_n^m)^{\frac{\gamma}{1-\beta}} (\tau_{nc}w_c)^{-\frac{1}{1-\beta}}$. The market clearing condition for low-skill workers is

$$L_{nc}^m \int_0^{\infty} \iota(\bar{a}) l(\bar{a}) d\bar{a} = L_{nc}^p.$$

That is

$$\kappa_{nc} \theta \Phi_n \tau_{nc}^{\theta} w_c^{\beta\theta} \int_0^{\infty} e^{-\Phi_n \tau_{nc}^{\theta} w_c^{\beta\theta} \bar{a}^{-\theta}} \bar{a}^{-\theta-1+\frac{1}{1-\beta}} d\bar{a} = \frac{L_{nc}^p}{L_{nc}^m}.$$

Or

$$\kappa_{nc} \theta \int_0^{\infty} e^{-y^{-\theta}} y^{-\theta-1+\frac{1}{1-\beta}} dy = \frac{L_{nc}^p}{L_{nc}^m} (\Phi_n \tau_{nc}^{\theta} w_c^{\beta\theta})^{-\frac{1}{\theta(1-\beta)}}.$$

The total output is given by (up to a constant)

$$\frac{(L_n^m)^\gamma}{\tau_{nc}}(\Phi_n \tau_{nc}^\theta w_c^{\beta\theta})^{\frac{1}{\theta}}(L_{nc}^p)^\beta (L_{nc}^m)^{(1-\beta)}.$$

We denote $S_{nc}(x, y) = (L_n^m)^\gamma(\Phi_n \tau_{nc}^\theta w_c^{\beta\theta})^{\frac{1}{\theta}} x^\beta y^{1-\beta}$, where $(L_n^m)^\gamma(\Phi_n \tau_{nc}^\theta w_c^{\beta\theta})^{\frac{1}{\theta}}$ is evaluated in the scenario without the internet.

K Equilibrium with Elastic Housing Supply

In the baseline model, we assume that housing supply is fixed. We relax this assumption and adopt the following form of elastic housing supply in each city n

$$H_n = \bar{H}_n p_n^\epsilon, \tag{66}$$

where $\bar{H}_n$ is a constant and ϵ is the supply elasticity. Then the equilibrium housing rent in city n will be revised to

$$p_n = \left[\frac{(1-\alpha)W_n}{\bar{H}_n}\right]^{\frac{1}{\epsilon+1}}. \tag{67}$$

The relative housing price between any two cities is given by (revised from (54))

$$\left(\frac{w_n}{w_c}\right)^{\frac{1}{1-\alpha}} = \frac{\left((1-\alpha)(\zeta({L_n^m}^{\gamma\theta}\Phi_n)^{\frac{1}{\theta(1-\beta)}} L_n^m + \sum_k \eta\left(T_k(\tau_{kn} w_n^\beta)^{-\theta}\right) \Phi_k^{\frac{1}{\theta(1-\beta)}-1} [L_k^m]^{\frac{\gamma}{1-\beta}+1})/\bar{H}_n\right)^{\frac{1}{\epsilon+1}}}{\left((1-\alpha)(\zeta({L_c^m}^{\gamma\theta}\Phi_c)^{\frac{1}{\theta(1-\beta)}} L_c^m + \sum_k \eta\left(T_k(\tau_{kc} w_c^\beta)^{-\theta}\right) \Phi_k^{\frac{1}{\theta(1-\beta)}-1} [L_k^m]^{\frac{\gamma}{1-\beta}+1})/\bar{H}_c\right)^{\frac{1}{\epsilon+1}}}. \tag{68}$$

$$\frac{\gamma}{1-\beta}\log\frac{L_n^m}{L_c^m} + \frac{1}{\theta(1-\beta)}\log\frac{\Phi_n}{\Phi_c} = (1-\alpha)\log\frac{p_n}{p_c} = \log\frac{w_n}{w_c} \tag{69}$$

If $\epsilon_n = \epsilon_c = \epsilon$, then

$$\begin{bmatrix} \zeta\Phi_1^{\frac{1}{\theta(1-\beta)}} + \eta T_1 \Delta_{11} w_1^{-\beta\theta} \Phi_1^{\frac{1}{\theta(1-\beta)}-1} & \eta T_2 \Delta_{21} w_1^{-\beta\theta} \Phi_2^{\frac{1}{\theta(1-\beta)}-1} & \dots & \eta T_N \Delta_{N1} w_1^{-\beta\theta} \Phi_N^{\frac{1}{\theta(1-\beta)}-1} \\ \eta T_1 \Delta_{12} w_2^{-\beta\theta} \Phi_1^{\frac{1}{\theta(1-\beta)}-1} & \zeta\Phi_2^{\frac{1}{\theta(1-\beta)}} + \eta T_2 \Delta_{22} w_2^{-\beta\theta} \Phi_2^{\frac{1}{\theta(1-\beta)}-1} & \dots & \eta T_N \Delta_{N2} w_2^{-\beta\theta} \Phi_N^{\frac{1}{\theta(1-\beta)}-1} \\ \vdots & \vdots & \ddots & \vdots \\ \eta T_1 \Delta_{1N} w_N^{-\beta\theta} \Phi_1^{\frac{1}{\theta(1-\beta)}-1} & \eta T_2 \Delta_{2N} w_N^{-\beta\theta} \Phi_2^{\frac{1}{\theta(1-\beta)}-1} & \dots & \zeta\Phi_N^{\frac{1}{\theta(1-\beta)}} + \eta T_N \Delta_{NN} w_N^{-\beta\theta} \Phi_N^{\frac{1}{\theta(1-\beta)}-1} \end{bmatrix}$$

$$\times \begin{bmatrix} w_1^{\frac{1-\beta}{\gamma}+1} \Phi_1^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\ w_2^{\frac{1-\beta}{\gamma}+1} \Phi_2^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\ \vdots \\ w_N^{\frac{1-\beta}{\gamma}+1} \Phi_N^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \end{bmatrix} = \begin{bmatrix} w_1^{\frac{1}{1-\alpha}(1+\epsilon)} H_1 \\ w_2^{\frac{1}{1-\alpha}(1+\epsilon)} H_2 \\ \vdots \\ w_N^{\frac{1}{1-\alpha}(1+\epsilon)} H_N \end{bmatrix}. \tag{70}$$

Multiply both sides by w_n^{aux}, where $aux \in \mathcal{R}$ is an auxiliary parameter, and we get

$$\begin{bmatrix} \zeta\Phi_1^{\frac{1}{\theta(1-\beta)}} w_1^{aux} + \eta T_1 \Delta_{11} w_1^{-\beta\theta} \Phi_1^{\frac{1}{\theta(1-\beta)}-1} w_1^{aux} & \dots & \eta T_N \Delta_{N1} w_1^{-\beta\theta} \Phi_N^{\frac{1}{\theta(1-\beta)}-1} w_1^{aux} \\ \eta T_1 \Delta_{12} w_2^{-\beta\theta} \Phi_1^{\frac{1}{\theta(1-\beta)}-1} w_2^{aux} & \dots & \eta T_N \Delta_{N2} w_2^{-\beta\theta} \Phi_N^{\frac{1}{\theta(1-\beta)}-1} w_2^{aux} \\ \vdots & \ddots & \vdots \\ \eta T_1 \Delta_{1N} w_N^{-\beta\theta} \Phi_1^{\frac{1}{\theta(1-\beta)}-1} w_N^{aux} & \dots & \zeta\Phi_N^{\frac{1}{\theta(1-\beta)}} w_N^{aux} + \eta T_N \Delta_{NN} w_N^{-\beta\theta} \Phi_N^{\frac{1}{\theta(1-\beta)}-1} w_N^{aux} \end{bmatrix}$$

$$\times \begin{bmatrix} w_1^{\frac{1-\beta}{\gamma}+1} \Phi_1^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\ w_2^{\frac{1-\beta}{\gamma}+1} \Phi_2^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\ \vdots \\ w_N^{\frac{1-\beta}{\gamma}+1} \Phi_N^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \end{bmatrix} = \begin{bmatrix} w_1^{\frac{1}{1-\alpha}(\mathbf{1}+\epsilon)+\mathbf{aux}} H_1 \\ w_2^{\frac{1}{1-\alpha}(\mathbf{1}+\epsilon)+\mathbf{aux}} H_2 \\ \vdots \\ w_N^{\frac{1}{1-\alpha}(\mathbf{1}+\epsilon)+\mathbf{aux}} H_N \end{bmatrix}. \tag{71}$$

We now have two additional parameters to calibrate. For the housing supply elasticity ϵ, we follow Giannone (2022) and set $\epsilon = 0.135$. We additionally obtain the housing parameter $\bar{H}_n$ by:

$$\frac{\bar{H}_n}{\bar{H}'_n} = \frac{W_n / w_n^{\frac{1}{1-\alpha}(1+\epsilon)}}{W'_n / w_n'^{\frac{1}{1-\alpha}(1+\epsilon)}}$$

We find that with elastic housing supply, the role of internet quality improvement in increasing welfare and skill redistribution is even bigger quantitatively. The welfare increase for managers becomes 3.89% and 3.67% for workers, respectively. Moreover, Table 12 shows that removing Internet imply a reduction of 0.0037 in the elasticity of high-skill employment share with respect to city size. This number is bigger than the baseline counterpart with inelastic housing supply. The intuition is straightforward: Elastic housing supply makes it easier for the labor market to adjust across regions.

Dependent variable:	Change in high-skill employment share with internet	
	(1)	(2)
City Size	0.0039***	0.0037***
	(0.0008)	(0.0011)
State fixed effects	No	Yes
Observations	722	722
R^2	0.049	0.075

$*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$

Table 12: High-skill Employment Share and City Size with Elastic Housing Supply

Notes: The dependent variable is the change in high-skill employment share after a more uniform internet quality improvement program. City size is measured by log(labor supply in 1980). Column (1) reports results using robust standard errors, and Column (2) with standard errors clustered by state.

L Amenity Spillover

We extend the baseline model to consider amenity spillovers. The extension is in the spirit of Diamond (2016). The utility function is revised to

$$c^{\alpha} h^{1-\alpha} A^{\zeta_s}$$

where $s = p, m$. The amenity supply is

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

The indifference conditions will then be changed to

$$\log w_n - (1-\alpha)\log p_n + \zeta^p \log A_n = \log w_c - (1-\alpha)\log p_c + \zeta^p \log A_c \tag{72}$$

and

$$\begin{aligned} &\frac{\gamma}{1-\beta}\log L_n^m + \frac{1}{\theta(1-\beta)}\log \Phi_n - (1-\alpha)\log p_n + \zeta^m \log A_n \\ &= \frac{\gamma}{1-\beta}\log L_c^m + \frac{1}{\theta(1-\beta)}\log \Phi_c - (1-\alpha)\log p_c + \zeta^m \log A_c. \end{aligned} \tag{73}$$

That is

$$\log w_n - (1-\alpha)\log p_n + \zeta^p\theta_a(\log L_n^m - \log L_n^p) = \log w_c - (1-\alpha)\log p_c + \zeta^p\theta_a(\log L_c^m - \log L_c^p) \tag{74}$$

and

$$\begin{aligned}&\frac{\gamma}{1-\beta}\log L_n^m + \frac{1}{\theta(1-\beta)}\log\Phi_n - (1-\alpha)\log p_n + \zeta^m\theta_a(\log L_n^m - \log L_n^p)\\ &= \frac{\gamma}{1-\beta}\log L_c^m + \frac{1}{\theta(1-\beta)}\log\Phi_c - (1-\alpha)\log p_c + \zeta^m\theta_a(\log L_c^m - \log L_c^p).\end{aligned} \tag{75}$$

Combining the above two to get

$$\begin{aligned}&\frac{\gamma}{1-\beta}\log L_n^m + \frac{1}{\theta(1-\beta)}\log\Phi_n - \log w_n + (\zeta^m\theta - \zeta^p\theta)(\log L_n^m - \log L_n^p)\\ &= \frac{\gamma}{1-\beta}\log L_c^m + \frac{1}{\theta_a(1-\beta)}\log\Phi_c - \log w_c + (\zeta^m\theta_a - \zeta^p\theta_a)(\log L_c^m - \log L_c^p).\end{aligned} \tag{76}$$

This implies that

$$L_n^{m\frac{\gamma}{1-\beta}+(\zeta^m-\zeta^p)\theta_a} \propto w_n\Phi_n^{-\frac{1}{\theta(1-\beta)}} L_n^{p(\zeta^m-\zeta^p)\theta_a}.$$

On the other hand,

$$\frac{p_n}{p_c} = \left(\frac{w_n}{w_c}\right)^{\frac{1}{1-\alpha}} \left(\frac{L_n^m}{L_c^m}\right)^{\frac{\zeta^p\theta_a}{1-\alpha}} \left(\frac{L_n^p}{L_c^p}\right)^{-\frac{\zeta^p\theta_a}{1-\alpha}}$$

Then we express the above in matrix form as

$$\begin{aligned}&\begin{bmatrix} \zeta\Phi_1^{\frac{1}{\theta(1-\beta)}} w_1^{aux} + \eta T_1\Delta_{11} w_1^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1} w_1^{aux} & \cdots & \eta T_N\Delta_{N1} w_1^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1} w_1^{aux} \\ \eta T_1\Delta_{12} w_2^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1} w_2^{aux} & \cdots & \eta T_N\Delta_{N2} w_2^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1} w_2^{aux} \\ \vdots & \ddots & \vdots \\ \eta T_1\Delta_{1N} w_N^{-\beta\theta}\Phi_1^{\frac{1}{\theta(1-\beta)}-1} w_N^{aux} & \cdots & \zeta\Phi_N^{\frac{1}{\theta(1-\beta)}} w_N^{aux} + \eta T_N\Delta_{NN} w_N^{-\beta\theta}\Phi_N^{\frac{1}{\theta(1-\beta)}-1} w_N^{aux} \end{bmatrix}\\ &\times \begin{bmatrix} w_1^{\frac{1-\beta}{\gamma}+1}\Phi_1^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\ w_2^{\frac{1-\beta}{\gamma}+1}\Phi_2^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \\ \vdots \\ w_N^{\frac{1-\beta}{\gamma}+1}\Phi_N^{-\frac{1}{\theta(1-\beta)}-\frac{1}{\theta\gamma}} \end{bmatrix} = \begin{bmatrix} w_1^{\frac{1}{1-\alpha}+aux} H_1 \left(\frac{L_1^m}{L_1^p}\right)^{\frac{\zeta^p\theta_a}{1-\alpha}} \\ w_2^{\frac{1}{1-\alpha}+aux} H_2 \left(\frac{L_2^m}{L_2^p}\right)^{\frac{\zeta^p\theta_a}{1-\alpha}} \\ \vdots \\ w_N^{\frac{1}{1-\alpha}+aux} H_N \left(\frac{L_N^m}{L_N^p}\right)^{\frac{\zeta^p\theta_a}{1-\alpha}} \end{bmatrix}.\end{aligned} \tag{77}$$

We make the following simplifications to illustrate how amenity spillover can shape the skill redistribution. Let $\zeta_p = 0$ and we introduce one more parameter $\theta_a\zeta_m > 0$ so that the endogenous amenities benefit the high-skilled more as in Diamond (2016). To investigate the role of amenity spillover, we set $\theta_a \zeta_m = 0.002$ and keep $\gamma = 0.010$ to re-compute how Internet connectivity changes the skill redistribution.[29] We find that the change in the elasticity of high-skill employment share becomes even higher (=0.0039) after Internet in 2013 is introduced. This result is intuitive since the endogenous amenities is assumed to benefit the high-skilled more thus promotes them to concentrate more in big cities.

M Fragmentation Sensitivity by Sectors

In the baseline model, there is only one production sector. We now do an extension to consider two sectors with differential fragmentation sensitivities. Suppose that sector 1 is fragmentation sensitive so that Internet quality improvement leads to lower iceberg cost in fragmentation. But in sector 2, the fragmentation cost cannot be reduced with the Internet and is set to be infinite so that inter-city joint production is not possible. We list the new equilibrium conditions below. Note that sector 1 communication cost between city n and c $\tau_{nc,1}$ is sensitive to the communication technology.

We add subscript $s = 1, 2$ to denote variables in different sectors. The labor market clearing conditions are given by

$$L^m = \sum_n L^m_{n,s} = \sum_{n,c} L^m_{nc,s}, \tag{78}$$

and

$$L^p = \sum_n L^p_{n,s} = \sum_{n,c} L^p_{nc,s} \tag{79}$$

for managers and production workers, respectively. The demand for production workers is given by

$$L^p_{nc,s} = \eta w_c^{-1} \left(T_{n,s} (\tau_{nc,s}/p^s w_c^{\beta})^{-\theta} \right) \Phi_{n,s}^{\frac{1}{\theta(1-\beta)}-1} [f(L^m_{n,s})]^{\frac{1}{1-\beta}} L^m_{n,s}. \tag{80}$$

We assume that the agglomeration force are at the origin city-sector level such that

$$\Phi_{n,s} = \sum_k T_{n,s} (\tau_{nk,s}/p^s w_k^{\beta})^{-\theta}.$$

[29] We keep elastic housing supply such that $\epsilon = 0.135$.

The final consumption bundle is assumed to be a Cobb-Douglas form

$$c = c_1^{\frac{1}{2}} c_2^{\frac{1}{2}}.$$

Normalizing the consumption bundle for tradables as 1

$$(2p^1)^{1/2}(2p^2)^{1/2} = 1.$$

The indifference conditions of managers and workers across different locations are

$$\log w_n - (1-\alpha)\log p_n = \log w_c - (1-\alpha)\log p_c.$$

$$\frac{\gamma}{1-\beta}\log L_{n,s}^m + \frac{1}{\theta(1-\beta)}\log \Phi_{n,s} - \log w_n = \frac{\gamma}{1-\beta}\log L_{c,s}^m + \frac{1}{\theta(1-\beta)}\log \Phi_{c,s} - \log w_c \tag{81}$$

Moreover, managers in the same city are indifferent in the sector they work

$$\frac{\gamma}{1-\beta}\log L_{n,s}^m + \frac{1}{\theta(1-\beta)}\log \Phi_{n,s} = \frac{\gamma}{1-\beta}\log L_{n,s'}^m + \frac{1}{\theta(1-\beta)}\log \Phi_{n,s'}.$$

The housing market clearing condition is

$$p_n = \frac{(1-\alpha)W_n}{H_n}$$

where W_n is the total income in city n

$$W_n = \sum_s \zeta(L_{n,s}^{m\gamma\theta}\Phi_{n,s})^{\frac{1}{\theta(1-\beta)}} L_{n,s}^m + \sum_{k,s}\eta(T_{k,s}(\tau_{kn,s}/p^s w_n^\beta)^{-\theta})\Phi_{k,s}^{\frac{1}{\theta(1-\beta)}-1}[L_{k,s}^m]^{\frac{\gamma}{1-\beta}+1}.$$

Goods market clearing conditions by sector are

$$\frac{1}{2}\sum_n W_n = \sum_{n,c} w_c L_{nc,s}^p/\beta$$

i.e.,

$$\frac{1}{2}\sum_n W_n = \sum_{n,c}\eta(T_{n,s}(\tau_{nc,s}/p^s w_c^\beta)^{-\theta})\Phi_{n,s}^{\frac{1}{\theta(1-\beta)}-1}[L_{n,s}^m]^{\frac{\gamma}{1-\beta}+1}/\beta$$

Therefore, for any two cities n and c, we get

$$\left(\frac{w_n}{w_c}\right)^{\frac{1}{1-\alpha}} = \frac{\sum_s \zeta(L_{n,s}^{m\gamma\theta}\Phi_{n,s})^{\frac{1}{\theta(1-\beta)}} L_{n,s}^m + \sum_{k,s} \eta(T_{k,s}(\tau_{kn,s}/p^s w_n^\beta)^{-\theta})\Phi_{k,s}^{\frac{1}{\theta(1-\beta)}-1}[L_{k,s}^m]^{\frac{\gamma}{1-\beta}+1}}{\sum_s \zeta(L_{c,s}^{m\gamma\theta}\Phi_{c,s})^{\frac{1}{\theta(1-\beta)}} L_{c,s}^m + \sum_{k,s} \eta(T_{k,s}(\tau_{kc,s}/p^s w_c^\beta)^{-\theta})\Phi_{k,s}^{\frac{1}{\theta(1-\beta)}-1}[L_{k,s}^m]^{\frac{\gamma}{1-\beta}+1}} \frac{H_c}{H_n}. \tag{82}$$

and

$$\frac{\gamma}{1-\beta}\log L_{n,s}^m + \frac{1}{\theta(1-\beta)}\log \Phi_{n,s} - \log w_n = \frac{\gamma}{1-\beta}\log L_{c,s}^m + \frac{1}{\theta(1-\beta)}\log \Phi_{c,s} - \log w_c, \tag{83}$$

where $s = 1, 2$.

Finally, the output produced by managers in city n and workers in city c of sector s is

$$L_{ncs}^m [L_{n,s}^m]^\gamma \frac{p^s}{\tau_{nc}} * \kappa_{ncs}^\beta * \theta * \int_0^\infty e^{-y^{-\theta}} y^{-\theta-1+\frac{1}{1-\beta}} dy * \Phi_{ns}^{-\frac{1}{\theta(1-\beta)}} [\tau_{ncs}^{-\theta}]^{\frac{1}{(1-\beta)\theta}} w_c^{\frac{\beta}{1-\beta}}$$

where $\kappa_{ncs} = \beta^{\frac{1}{1-\beta}} [L_{n,s}^m]^{\frac{\gamma}{1-\beta}} [\frac{\tau_{ncs} w_c}{p^s}]^{-\frac{1}{1-\beta}}$.

We solve a two-city two-sector case numerically and show the spatial dispersion of high-skill employment shares for the two sectors, respectively in Figure 10. We can find that when there is Internet quality increases (ICT openness increases for sector 1), the high-skill employment shares of sector 1 diverge in the two cities as reflected in the standard deviations. But this is not the case for sector 2, which does not have a tendency to fragment even with ICT improvement.

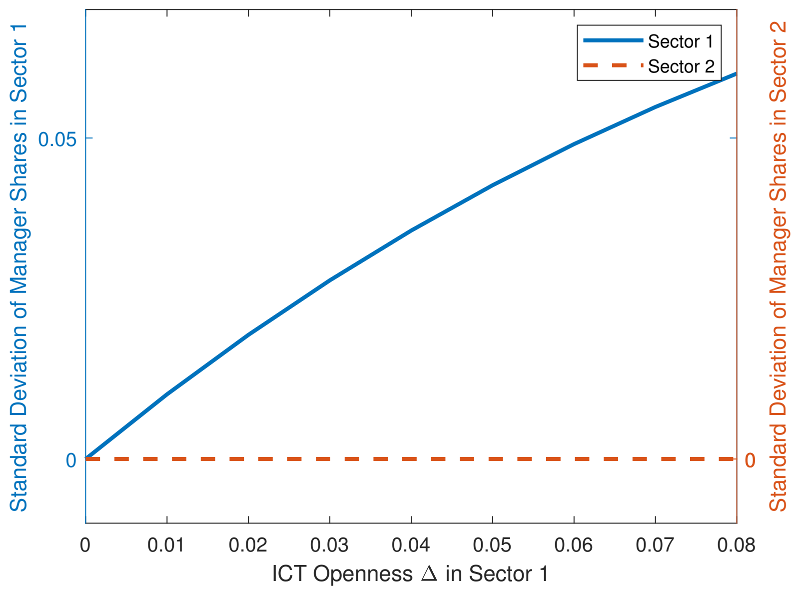


Figure 10: Two-city Two-Sector Equilibrium: Spatial Dispersion of High-skill Employment Share in Two Sectors

Notes: In this simulation, we set $\theta = 5$, city 1 technology $T_1^{\frac{1}{\theta}} = 1.5$ for both sectors, city 2 technology $T_2^{\frac{1}{\theta}} = 1.0$ for both sectors, $L_m = 1$, $L_p = 2$, $\gamma = 0.1$, $\alpha = 0.4$, $\beta = 0.4$, $H_1 = H_2 = 1.0$. Sector 1 is fragmentation sensitive while sector 2 is not. The horizontal axis shows the extent of ICT openness $\triangle \equiv \tau^{-\theta}$ in sector 1.

N Additional Figures and Tables

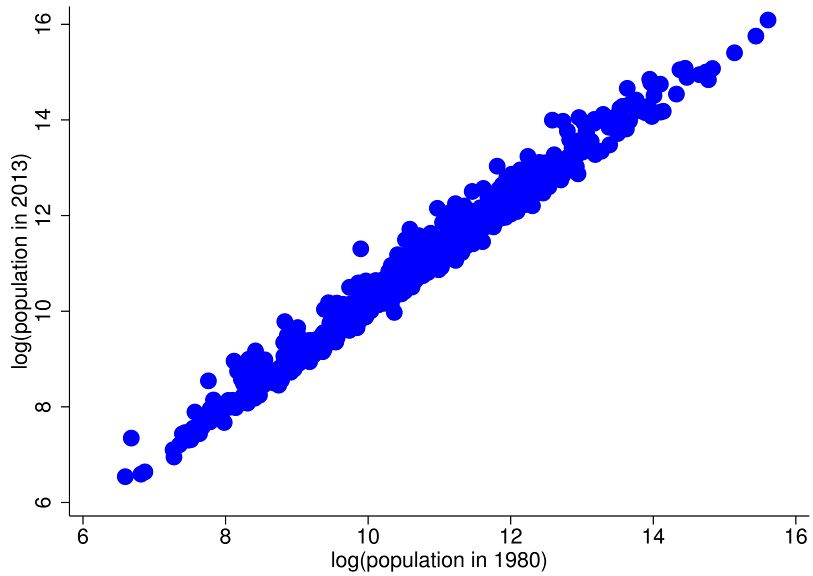


Figure 11: log(population in 2013) against log(population in 1980) across CZs

Notes: Each dot represents a commuting zone. The linear correlation between log(labor supply in 2013) and log(labor supply in 1980) is 0.99.

Code	Industry Description	Fort Index	Δ KM Index
100	Meat products	0.111	-0.013
101	Dairy products	0.142	0.048
102	Canned, frozen, and preserved fruits and vegetables	0.194	0.097
110	Grain mill products	0.171	0.156
111	Bakery products	0.082	-0.020
112	Sugar and confectionery products	0.140	0.046
120	Beverage industries	0.189	0.342
121	Misc. food preparations and kindred products	0.198	0.087
130	Tobacco manufactures	0.267	0.326
132	Knitting mills	0.292	0.086
140	Dyeing and finishing textiles, except wool and knit goods	0.269	0.246
141	Carpets and rugs	0.253	-0.019
142	Yarn, thread, and fabric mills	0.274	0.221
150	Miscellaneous textile mill products	0.212	N/A
151	Apparel and accessories, except knit	0.272	0.098
152	Miscellaneous fabricated textile products	0.228	0.125
160	Pulp, paper, and paperboard mills	0.239	0.146
161	Miscellaneous paper and pulp products	0.362	0.148
162	Paperboard containers and boxes	0.362	0.042
172	Printing, publishing, and allied industries, except newspapers	0.322	0.108
180	Plastics, synthetics, and resins	0.263	0.030
181	Drugs	0.324	1.070
182	Soaps and cosmetics	0.325	0.328
190	Paints, varnishes, and related products	0.223	0.321
191	Agricultural chemicals	0.126	0.291
192	Industrial and miscellaneous chemicals	0.196	0.373
200	Petroleum refining	0.155	0.110
201	Miscellaneous petroleum and coal products	0.155	0.453
210	Tires and inner tubes	0.257	0.122
211	Other rubber products, and plastics footwear and belting	0.257	0.123
212	Miscellaneous plastics products	0.231	0.089
222	Leather products, except footwear	0.276	0.181
231	Sawmills, planing mills, and millwork	0.117	0.120
232	Wood buildings and mobile homes	0.224	0.146
241	Miscellaneous wood products	0.239	N/A
242	Furniture and fixtures	0.650	0.097

Table 13: Change in KM Segregation Index and Fort (2017) Fragmentation Index

Code	Industry Description	Fort Index	Δ KM Index
250	Glass and glass products	0.238	0.156
251	Cement, concrete, gypsum, and plaster products	0.118	0.098
252	Structural clay products	0.192	0.195
261	Pottery and related products	0.192	0.191
262	Misc. nonmetallic mineral and stone products	0.137	0.097
270	Blast furnaces, steelworks, rolling and finishing mills	0.332	0.120
271	Iron and steel foundries	0.362	0.075
272	Primary aluminum industries	0.270	0.063
280	Other primary metal industries	0.272	0.096
281	Cutlery, handtools, and general hardware	0.378	0.244
282	Fabricated structural metal products	0.302	0.072
290	Screw machine products	0.352	N/A
291	Metal forgings and stampings	0.404	0.119
300	Miscellaneous fabricated metal products	0.318	0.164
310	Engines and turbines	0.501	0.450
311	Farm machinery and equipment	0.414	0.237
312	Construction and material handling machines	0.414	0.301
320	Metalworking machinery	0.409	0.233
321	Office and accounting machines	0.400	N/A
322	Computers and related equipment	0.501	1.833
331	Machinery, except electrical, n.e.c.	0.375	0.218
340	Household appliances	0.374	0.215
341	Radio, TV, and communication equipment	0.489	1.192
342	Electrical machinery, equipment, and supplies, n.e.c.	0.372	0.687
351	Motor vehicles and motor vehicle equipment	0.418	0.196
352	Aircraft and parts	0.500	0.494
360	Ship and boat building and repairing	0.228	0.163
361	Railroad locomotives and equipment	0.304	0.174
362	Guided missiles, space vehicles, and parts	0.500	0.845
370	Cycles and miscellaneous transportation equipment	0.466	0.337
371	Scientific and controlling instruments	0.467	0.793
372	Medical, dental, and optical instruments and supplies	0.278	0.440
381	Watches, clocks, and clockwork operated devices	0.467	N/A
390	Toys, amusement, and sporting goods	0.300	0.369
391	Miscellaneous manufacturing industries	0.300	0.194
610	Retail bakeries	0.082	0.088

Table 13 (continued). Change in KM Segregation Index and Fort (2017) Fragmentation Index

Notes: This table displays the change in the Kremer-Maskin (KM) skill segregation index between 1980 and 2013 and the Fort (2017) fragmentation index in each industry (based on 1990 Census industry classification). We use industry concordance between census industry classification and NAICS 4-digit available from the U.S. Census. When a Census industry corresponds to multiple NAICS industries, we calculate the simple average of fragmentation indices of the several NAICS industries.

Dependent variable: Change in high-skill employment share						
	(1)	(2)	(3)	(4)	(5)	(6)
high skilled defined by	67% cutoff		80% cutoff		college and above	
log(labor supply in 1980)	0.001	0.002**	0.005***	0.006***	0.004***	0.004***
	(0.001)	(0.0007)	(0.0006)	(0.0007)	(0.0007)	(0.0006)
state fixed effects	No	Yes	No	Yes	No	Yes
Observations	722	722	722	722	722	722
R^2	0.003	0.298	0.112	0.372	0.119	0.355

$* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$

Table 14: Change in High-Skill Employment Share and City Size: Robustness Checks

Notes: Columns (1)-(2) define the high skilled as occupations whose rank is above 67% of all occupations in 1980. Columns (3)-(4) define the high skilled as occupations whose rank is above 80% of all occupations in 1980. Columns (5)-(6) define the high skilled as workers who have a college education or above. Columns (1), (3), and (5) leave out the state fixed effect and report the robust standard errors. Columns (2), (4), and (6) use the state fixed effect and report standard errors clustered at the state level.

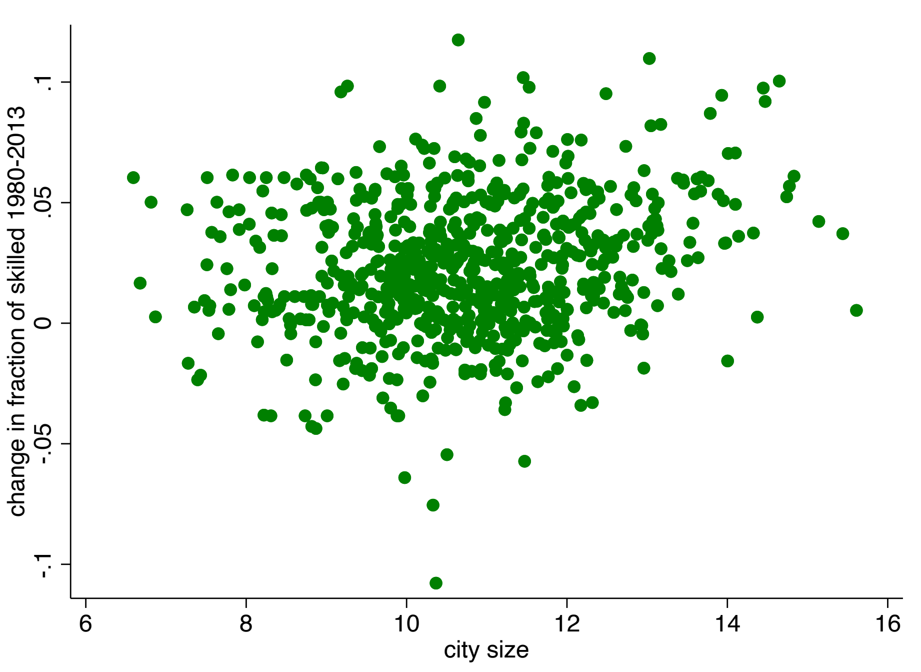


Figure 12: Change in High-skill Employment Share with Respect to City Sizes

Notes: This figure displays the change in the skilled share from 1980 to 2013 against log of 1980 labor supply (raw data). High skill is defined as occupation rank above 75% using the 1980 mean of log hourly wage.

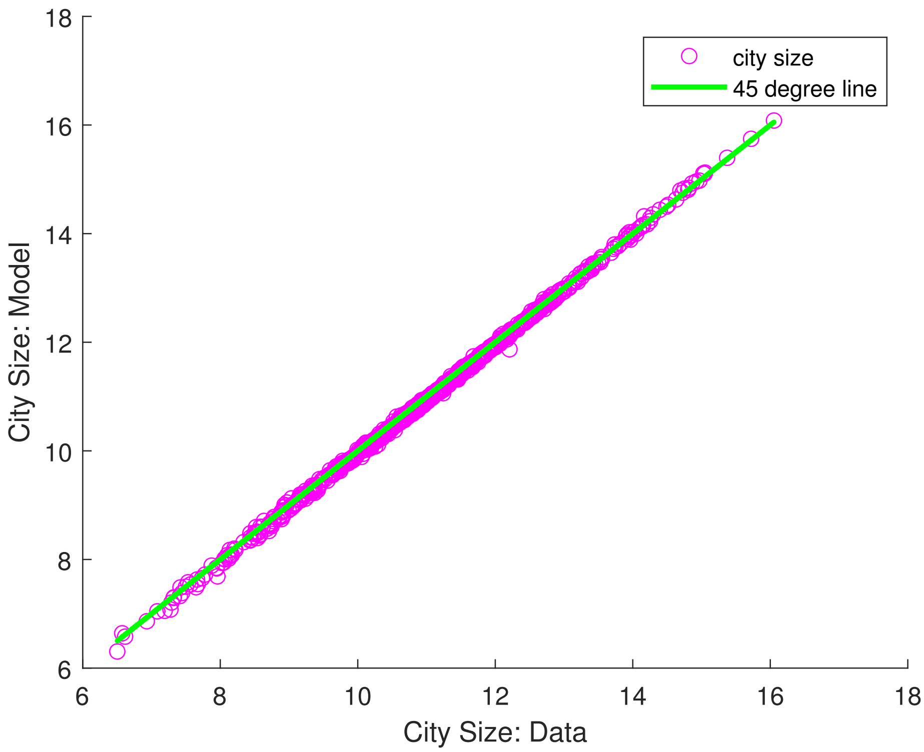


Figure 13: City Size in 2013: Model v.s. Data

Notes: We solve our model using the calibrated parameter values. We then calculate the model-implied equilibrium city sizes for each commuting zone and plot them against the actual log(labor supply in 2013).

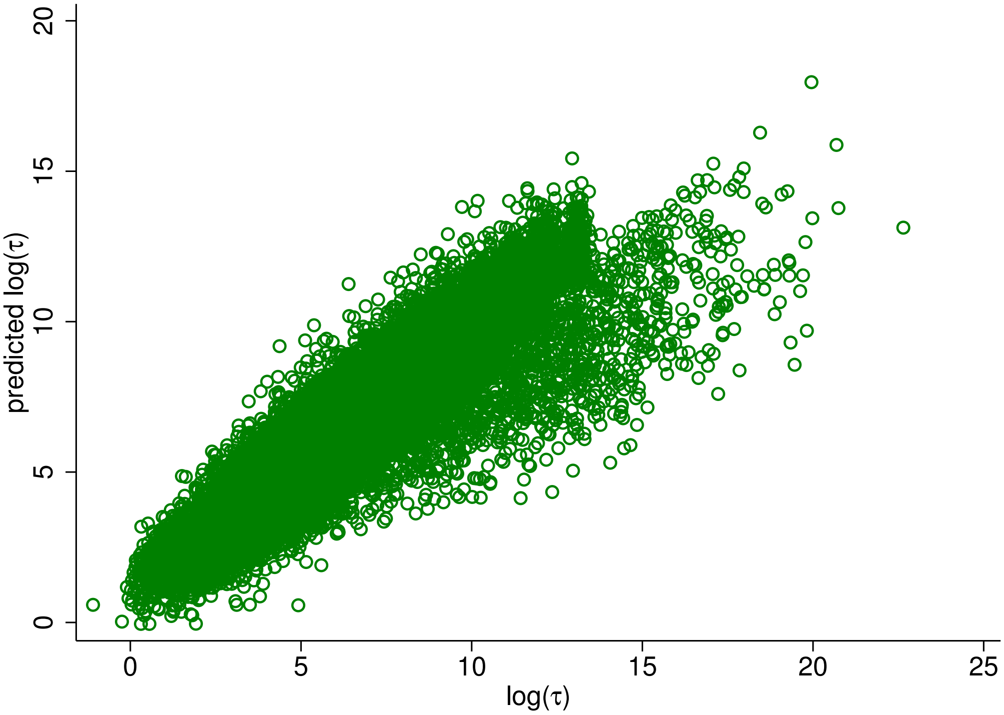


Figure 14: Fragmentation Cost Predicted by the Gravity Equation

Notes: This graph shows the predicted log(fragmentation cost τ) implied by the gravity equation test with the OLS estimation (controls are included) in Section 5. The x-axis is the data, and the y-axis is the predicted value.