The Geography of Innovation in the United States

Weiliang Tan*
Cornell University

January 21, 2025

JOB MARKET PAPER
Click here for the latest version

Abstract

A defining trend in U.S. innovation is its increasing geographic concentration, exemplified by the growth of high-tech clusters like Silicon Valley. What factors drive this increasing spatial concentration, and what are its implications for regional and aggregate growth? Using comprehensive data on patents, firms, and inventors from 1976 to 2018, I find that innovation became more concentrated in high-skill cities only after 1990, with the sudden rise of information and communication technologies (ICT) playing two distinct roles in this process. First, there was a compositional shift in innovation towards ICT, which is colocated with ICT production and concentrated in high-skill cities. Second, firms that were initially concentrated in high-skill cities produced more non-ICT patents likely due to spillovers from ICT innovation and ICT-enabled reductions in communication costs, which allowed these firms to expand production to lower-cost regions and enhanced the profitability of new ideas. Worker migration to high-skill cities amplified the effects of these mechanisms, intensifying the spatial concentration of innovation. To better understand the mechanics of innovation across space and its consequences for macroeconomic growth, I develop a model of spatial growth with endogenous and directed innovation, technology diffusion, and dynamic worker mobility with frictions. The model provides an analytical characterization of the *spatial direction of innovation* on the transition path and how its steady-state distribution across space determines long-run aggregate growth.

Keywords: spatial innovation; trade and growth; agglomeration economies; information and communication technologies

*Email: wt289@cornell.edu; weiliang.tan@yale.edu. I am extremely grateful to my advisors Ivan Rudik, Samuel Kortum, Peter Schott, and Nancy Chau for valuable advice. I am indebted to Professor Rudik for all his support and to Professor Kortum for hosting me at Yale since Spring 2023 and all his guidance. I thank Andy Bernard, Michael Peters, Costas Arkolakis, Lorenzo Caliendo, Shanjun Li, Ezra Oberfield, Bernardo Ribeiro, Sabrina Peng, and Zebang Xu for many helpful conversations, and Julieta Caunedo and Cecilia Fieler for helpful comments.

1 Introduction

The central questions in spatial economics are what drives the geographic distribution of economic activity and what are its local and aggregate consequences. This paper explores these questions in the context of the rising spatial concentration of U.S. innovation, a striking trend in recent decades exemplified by the growth of high-tech clusters like Silicon Valley. The agglomeration benefits of these clusters have been widely discussed in urban economics and the popular press. Most notably, Moretti (2021) finds that individual inventors are more productive when located in regions with a high density of inventors, with agglomeration spillovers resulting in greater aggregate innovation than if inventors were evenly distributed across space. More fundamentally, however, little is known formally about why the growth of these high-tech clusters has been a relatively recent phenomenon in the United States. Do shifts in industry composition, firm entry, worker sorting by skill, or initial conditions explain the rise of high-tech clusters? Understanding the root causes driving the rising spatial concentration of innovation is crucial for developing effective placed-based policies that might amplify these mechanisms and promote local growth. Additionally, the macroeconomic consequences of these high-tech clusters remain unclear. Do they boost the production of ideas that are adopted nationwide, thereby enhancing aggregate growth, or do these ideas remain confined within the high-tech clusters, potentially stifling broader economic development?

To answer these questions, it is essential first to carefully examine when and where the growth of high-tech clusters occurred. I begin by leveraging comprehensive data on patents and inventors to develop new measures that track how the geography of innovation in the United States has evolved over time. These measures reveal that the spatial concentration of patents remained relatively constant between 1976 and 1990 but increased dramatically from 1990 to 2018. Specifically, the Gini coefficient of patents per capita increased from approximately 0.36 in 1990 to 0.50 in 2018 – a change four times greater than the rise in income inequality in the US over the same period. Alternative measures of spatial concentration – such as the share of patents produced in the top 10 or 15 regions, the Herfindahl-Hirschman Index, and the Ellison-Glaeser measure – exhibit similar trends. Notably, this rising concentration primarily occurred in high-skill cities rather than in densely populated or large ones. The annual elasticity of patents per capita with respect to the Commuting Zone's (CZ) 1990 college ratio increased from about 1 in 1990 to 2 in 2018.

These facts suggest that a significant shock around 1990 triggered the growth of high-tech clusters. I find that the rapid rise of information and communication technologies (ICT) played two distinct roles in this process. First, there was a compositional shift in innovation toward ICT, with the share of patents in ICT rising from approximately 8.5% in 1990 to 33% in 2018. Notably, ICT patents are more concentrated in high-skill cities: both the Gini coefficient of ICT patents per capita and the elasticity of CZ ICT patents per capita with respect to the 1990 college ratio are nearly double those of patents in other fields. This greater concentration is likely driven by the *colocation* of ICT innovation with ICT production – comprising Software Publishers; Telecommunications; and Data Processing, Hosting, and Related Services – which was already concentrated in high-skill cities before 1990, presumably to access high-skill labor. A simple decomposition shows that the rising share of ICT patents accounts for 53% of the overall increase in the concentration of

innovation in high-skill cities from 1990 to 2018. Additionally, ICT production – and, consequently, ICT innovation – became even more concentrated in high-skill cities after 1990, accounting for an additional 13% of the overall increase.

Second, firms initially concentrated in high-skill cities produced significantly more non-ICT patents after 1990, accounting for the remaining third of the overall increase in the spatial concentration of innovation in high-skill cities from 1990 to 2018. In particular, a one-unit increase in firms' initial elasticity of patents per capita with respect to the 1990 college ratio between 1988 and 1990 is associated with a 20ppt greater increase in non-ICT patents in 2000 relative to 1990. I provide circumstantial evidence that the ICT shock from 1990 explains this compositional shift in non-ICT patents across firms through two mechanisms: (i) *spillovers* from ICT to non-ICT innovation, and (ii) the spatial expansion of these firms to lower-cost locations, which likely increased the profitability of new ideas – a phenomenon I term the *asymmetric scale effect*. Specifically, after controlling for firms' ICT patents, the greater increase in non-ICT patents among firms initially concentrated in high-skill cities is reduced by half. Although other shocks in the 1990s may have disproportionately benefited these firms, driving increases in both ICT and non-ICT patents, a more plausible explanation is that these firms gained from spillovers from ICT to non-ICT innovation following the ICT shock. Additionally, I find that these firms disproportionately expanded to lower-cost locations: a one-unit increase in a firm's elasticity of employment per capita with respect to the 1990 college ratio between 1988 and 1990 is associated with a 4ppt greater increase in the number of CZs within the firm's establishments in 2000 relative to 1990. More importantly, this expansion corresponds to a significantly larger decrease of 0.2 in the firm's elasticity of employment per capita with respect to the 1990 college ratio in 2000 relative to 1990. These estimates indicate that firms initially concentrated in high-skill cities were more likely to expand to lower-cost regions after 1990 but not before, providing evidence of the *asymmetric scale effect*: a novel productivity advantage of high-skill cities following the ICT shock.

My empirical findings suggest that the ICT shock from 1990 drove the rising spatial concentration of innovation in high-skill cities through three primary mechanisms: a compositional shift in innovation towards ICT, which is *colocated* with ICT production in high-skill cities; *spillovers* from ICT to non-ICT innovation; and the spatial expansion of firms initially concentrated in high-skill cities to lower-cost locations, i.e. the *asymmetric scale effect*. Worker migration to high-skill cities during this period amplified the impacts of these mechanisms on the spatial concentration of innovation. To formalize how these mechanisms shape the geography of innovation, I *introduce endogenous and directed innovation* into existing quantitative spatial models. To achieve this, I develop a model of spatial growth with endogenous innovation, technology diffusion, and dynamic worker mobility with frictions. Building on techniques from Eaton and Kortum (2001) and Lind and Ramondo (2024), I first integrate endogenous innovation and technology diffusion at the level of *individual ideas*, the fundamental unit in the Eaton-Kortum structure (Eaton and Kortum, 2024). Consequently, my model characterizes the degree of *colocation* between innovation and production and explains the *asymmetric scale effect*. I then introduce dynamic worker mobility with frictions, allowing workers to relocate in response to wage differences across regions and sectors, as well as between production and research. Worker mobility is driven by the two key mechanisms outlined above, in addition to *spillovers* from

ICT to non-ICT innovation, and shapes the spatial direction of innovation following the ICT shock.

My model features two sectors, ICT and non-ICT. In each region and sector, innovation workers discover new ideas through a Poisson process, where the arrival rate is influenced by various factors. Once an idea is discovered in a particular region, its diffusion to all regions is governed by independent Poisson processes with varying bilateral diffusion speeds. I capture the ICT shock through two exogenous components: (i) a one-time increase in the national productivity of ICT innovation, reflected in the Poisson arrival rate of new ideas; and (ii) a decrease in the national component of bilateral diffusion speeds, representing reduced communication costs. Each idea corresponds to the production of a specific good from the unit interval and has two additional attributes: stochastic quality, drawn from a Pareto distribution upon discovery, and stochastic applicability, drawn from a separate Pareto distribution upon arrival in a different region. The productivity of an idea in producing its corresponding good is determined by the product of its quality and applicability. This microfounded structure generates a multivariate productivity distribution for goods at each instant, where the marginal distribution in each region is Fréchet à la Eaton and Kortum (2002), but with an endogenous scale parameter determined by the past history of innovation in all regions and diffusion speeds in all region-pairs. Notably, the equilibrium trilateral trade shares are determined by the product of idea adoption and idea market shares, capturing two key mechanisms documented in my empirical findings: the idea adoption shares determine the degree of *colocation* between innovation and production, and the idea market shares explain the *asymmetric scale effect*, whereby a uniform increase in bilateral diffusion speeds disproportionately enhances the market access of ideas discovered in high-skill cities.

To enable these two mechanisms, along with *spillovers* from ICT to non-ICT innovation, to shape the geography of innovation, I introduce *endogenous and directed innovation* through imperfect competition – generating profits from production and, consequently, innovation – and worker mobility. In my model, there is a unit continuum of firms in each market (defined by region and sector), each employing innovation workers and owning their ideas. Each firm is thus a collection of ideas. These firms engage in Bertrand competition à la Bernard et al. (2003), where the lowest cost producer of each good claims the entire market for that good, charging the highest markup that deters any competitor from entering. This market structure generates an endogenous markup distribution across ideas, driven by their stochastic quality, while aggregate profits from the sale of all goods in any destination market is a constant share of total income or expenditure in that market. The expected value of an individual idea in any destination market is determined by the product of its market share and the total profits earned from the sale of all goods in that market. Consequently, wages from innovation are determined by the sum of the expected values of an idea across all destination markets, multiplied by the Poisson arrival rate in the region where innovation occurs. Given wages from production and innovation, workers make dynamic decisions about moving across regions and sectors, as well as between production and innovation. Along the transition path, the ratio of real wages from innovation across regions characterizes the incentives for worker mobility and, consequently, the spatial direction of innovation. In my context, I assume that the Poisson arrival rate of new ideas depends on the benefits of *colocation* with production, *spillovers* from ICT to non-ICT innovation, and the number of innovation workers in that region-sector (capturing the roles of the *asymmetric scale effect* and *worker migration*). Consequently, these

endogenous mechanisms influence innovation worker wages and hence worker mobility patterns, shaping the geography of innovation following the ICT shock, in line with my empirical findings.

An additional advantage of my model is its ability to characterize how the geography of innovation affects aggregate growth and shapes the welfare impact of the ICT shock, thereby establishing a direct connection between quantitative trade and spatial models with innovation and endogenous growth models in macroeconomics. Along the balanced growth path, prices in all regions fall at the same aggregate rate, mirroring macroeconomic growth models. Regional wages differ but remain constant over time, determined by the spatial distribution of workers, trade, and technology diffusion like in quantitative trade and spatial models. However, unlike macroeconomic growth models, the aggregate growth rate of prices in my model is endogenously determined by the spatial distribution of innovation rates. In contrast to trade and spatial models, a temporary shock to fundamentals in my model impacts not only the steady-state distribution of trade shares and nominal wages but also the long-run growth rate of the economy through falling prices. Specifically, I use the characterizations of the balanced growth and transition paths to *analytically decompose the welfare impact* of the ICT shock – or any other shock to economic fundamentals – into its *transitory and long-run growth components*.

Contributions to the Literature

This paper makes two central contributions across different literatures. First, I leverage comprehensive data on patents, inventors, and firms to examine when, where, and why innovation became increasingly spatially concentrated in the United States. While the empirical literature in innovation and urban economics has documented the rise of high-tech clusters in recent decades (e.g. Feldman and Kogler, 2010; Andrews and Whalley, 2021) and highlighted the benefits these clusters offer – such as enhancing inventor productivity and connecting innovation with academic science (e.g. Moretti, 2021; Bikard and Marx, 2020) – the more enduring and challenging question of what drives the growth of these high-tech clusters remains unanswered. I identify the growth of high-tech clusters from 1990 onward as having primarily occurred in high-skill cities, with the ICT shock likely driving this trend through three distinct mechanisms: *colocation*, *spillovers*, and the *asymmetric scale effect*. The first mechanism – the colocation of ICT innovation and production – is a sector-specific extension of the broader colocation of innovation and production documented within the United States by Fort et al. (2020) and across countries by Liu (2024). Specifically, Fort et al. (2020) show that firms and firm-regions with both innovation and production plants generate more patents than those without. I build on their findings by highlighting the role of ICT plants versus a broader set of innovation plants that includes R&D facilities. Additionally, I examine the intensive margin of colocation through employment shares and investigate how colocation influences the aggregate spatial concentration of innovation. The third mechanism highlights the geography of firm spatial expansion and its aggregate consequences on the spatial concentration of innovation. In particular, I present the first evidence that firms initially concentrated in high-skill cities expanded production to lower-cost regions significantly more than other firms after 1990, but not before. This extends the empirical firm network literature, which documents the spatial expansion of firms (e.g. Hsieh and Rossi-Hansberg, 2021; Kleinman, 2022; Jiang, 2023) but largely

abstracts from the geographic dimensions of this expansion.

Second, I introduce *endogenous and directed innovation in a spatial setting*, thereby contributing to the trade, spatial, and macroeconomics literatures. To achieve this, I develop a model of spatial growth with two key ingredients. The first ingredient integrates endogenous innovation with technology diffusion at the level of *individual ideas*, advancing the quantitative trade and innovation literature. Specifically, unlike Eaton and Kortum (2001) and Lind and Ramondo (2023, 2024), innovation in my framework is fully endogenous and depends not only on equilibrium trade but also the entire technology diffusion network. Consequently, I characterize the degree of *colocation* between innovation and production and explain the *asymmetric scale effect*. These equilibrium results correspond to the spatial mechanics of innovation documented in my empirical findings and cannot be derived in existing trade and growth models, where innovation is either highly stylized or modeled as independent of technology diffusion. Most notably, Buera and Oberfield (2020) model technology diffusion as the transfer of knowledge from existing goods to the creation of new ones, such that the technology diffusion network does not impact profits from innovation. Somale (2021) incorporates endogenous innovation but excludes technology diffusion. Two recent papers integrate trade, innovation, and diffusion, albeit with simplifying assumptions. Cai et al. (2021) assumes perfect substitutability of ideas diffused from different locations, resulting in a scenario where small changes in relative wages cause large shifts in idea adoption and trade shares – a feature Eaton and Kortum term "the problem of flats". Meanwhile, Xiang (2023) assumes instantaneous diffusion, limiting the model's ability to capture the spatially heterogeneous effects of a uniform increase in bilateral diffusion speeds on idea market access.

The second ingredient of my model is dynamic worker mobility with frictions across regions and sectors as well as between production and research. This contrasts with the existing quantitative trade and innovation literature, which typically assumes perfect mobility restricted solely to movements between production and research. With both ingredients, my model is able to characterize both the balanced growth path – the primary focus of this literature – and the transition path in response to shocks to economic fundamentals. Thus, my work aligns with the class of quantitative dynamic spatial models that incorporate worker migration (Caliendo et al., 2019), capital accumulation (Kleinman et al., 2023), and knowledge diffusion (Cai et al., 2022), all of which feature rich transition dynamics. I extend this class of models by introducing *endogenous and directed innovation* and by *decomposing the welfare effects of shocks to fundamentals into their transitory and long-run growth components*. Desmet et al. (2018) offers perhaps the only workhorse quantitative spatial model featuring endogenous innovation. However, they model endogenous innovation under perfect competition, where incentives to innovate arise from land rents. In contrast, my model fully microfounds innovation and technology diffusion *within* the Eaton-Kortum framework. Consequently, the spatial and sectoral directions of innovation on the transition path in my model depend not only on local population but also on equilibrium trade and technology diffusion networks. My analytical characterization of the spatial direction of innovation also advances the endogenous growth literature in macroeconomics, which has exclusively focused on the sectoral direction of technological change (e.g. Acemoglu, 1998, 2002, 2007) while abstracting from spatial considerations.

2 An Empirical Examination of the Rising Spatial Concentration of Innovation in the United States: When, Where, and Why

One of the most salient facts in the empirics of innovation is its rising spatial concentration in the United States, exemplified by the growth of Silicon Valley-like clusters. The fundamental drivers of this trend have, however, remained elusive despite more than two decades of research on the geography of innovation. Understanding why innovation became increasingly concentrated in space requires careful consideration of when and where it happened. In this section, I use the universe of patents, inventors, and patenting firms from 1976-2018 to: (i) develop new measures of the geography of U.S. innovation; (ii) document trends in the spatial concentration of innovation and its underlying geography, and; (iii) provide suggestive evidence on the key mechanisms driving these trends. I present these as nine facts after discussing my data sources.

Data

I obtain the universe of patents produced between 1976 and 2022 from PatentsView, supplemented with bulk files from the US Patent and Trademark Office (USPTO). My sample includes all 3.70 million utility patents where at least one inventor lists a US address. Each patent contains extensive information, including inventor addresses, which typically reflect their home city and state. Using the Google Maps API, I geocode these addresses and map them to various spatial resolutions within the United States. I use these inventor locations to calculate annual patent counts for each region and derive measures of the spatial concentration of innovation across regions over time. Additionally, every patent is assigned a unique primary Cooperative Patent Classification (CPC) technology class. To analyze the role of compositional changes in innovative activity across fields, I map these classes to broader technology fields and subfields by adapting the field classification methodology developed by the World Intellectual Property Organization. Most patents also have one or more assignees, typically US firms, that hold ownership of the intellectual property. To examine the roles of compositional changes across firms and firm spatial expansion, I link these patent assignees to the universe of firms in the restricted US Census Longitudinal Business Database (LBD) using crosswalks provided by Kerr and Fu (2008) and Dreisigmeyer et al. (2018). Within firms, I assign inventors to regions in two alternative ways: the commuting zone (CZ) of their home city, and the nearest CZ where the firm has an establishment following Fort et al. (2020). I supplement the US Census LBD data with data on public firms and their establishments from 1990-2018 sourced from Dun and Bradstreet's National Establishment Time Series (NETS) Database. In my empirical analysis, I refer to firms interchangeably as patent assignees, public firms from the NETS dataset, and patenting firms from the LBD. See Appendix A for more details.

List of Facts

My facts highlight the key mechanisms driving the rising geographical concentration of innovation in recent decades and are as follows: (1) innovation became more concentrated in high-skill cities after 1990; (2) the rapid rise of information and communication technologies (ICT) occurred primarily after 1990; (3) ICT innovation is more concentrated in high-skill cities than other fields; (4) ICT innovation is *colocated*

with ICT production, both of which became more concentrated in high-skill cities after 1990; (5) non-ICT innovation also became more concentrated in high-skill cities after 1990; (6) firms initially concentrated in high-skill cities drove non-ICT patent growth after 1990; (7) these firms likely benefited disproportionately from *spillovers* of ICT innovation; (8) these firms increasingly and disproportionately expanded to low-skill regions after 1990 (the *asymmetric scale effect*); and (9) both low and high-skill workers migrated to high-skill cities from 1990.

Specifically, I attribute the rising geographical concentration of innovation in high-skill cities since 1990 (Fact 1) to the effects of the ICT shock (Fact 2) on ICT innovation (Facts 3 and 4) through *colocation* and on non-ICT innovation (Facts 5 to 8) through *spillovers* and the *asymmetric scale effect*. Worker migration to high-skill cities (Fact 9) amplifies the impact of the ICT shock.

Fact 1: Innovation became more concentrated in high-skill cities after 1990

I begin by using the locational Gini index (Krugman, 1991) to measure the aggregate spatial concentration of innovation annually from 1976 to 2018. For each commuting zone (CZ), I calculate its share of total patents and share of total population in the US. I then rank CZs by their patent-to-population share ratio and plot the cumulative sum of patent shares against the cumulative sum of population shares to construct the locational Gini curve. The locational Gini index, proportional to the area between this curve and the 45-degree line, measures the concentration of patents across CZs relative to population, with each CZ weighted by its population share. A value of zero indicates perfect equality while a value of one reflects perfect inequality[1].

Using this index, the solid black line in the left panel of Figure 1 shows that the spatial concentration of patents remained approximately constant from 1976 to 1990 but increased significantly from 1990 to 2018. The dotted grey line depicts the trend in the spatial concentration of inventors, computed annually by substituting patent shares with inventor shares in the locational Gini index. This trend closely mirrors that for patents, with the gap between the two from 1995 to 2018 potentially reflecting agglomeration economies in innovation, as documented by Moretti (2021)[2]. The dot-dashed blue line illustrates trends in the locational Gini index of patents relative to college-educated workers, computed annually by replacing population shares with college-educated worker shares in the locational Gini index. Relative to college-educated workers, the spatial concentration of patents remained approximately constant in the 1980s and rose significantly after 1990. This finding suggests that the rising spatial concentration of innovation cannot be attributed to the

[1]Formally, the locational Gini index G of patents x with respect to population L across all the N CZs is equivalent to half of the weighted mean absolute deviation of patents across all CZ-pairs o, d:

$$G = \frac{1}{2\mu} \sum_{o=1}^{N} \sum_{d=1}^{N} \frac{L_o}{L} \frac{L_d}{L} |x_o - x_d|, \qquad \mu = \sum_{o=1}^{N} \frac{L_o}{L} x_o.$$

[2]Specifically, Moretti (2021) shows that inventors produce more patents when located in regions with more inventors. This finding implies that the spatial concentration of patents exceeds that of inventors.

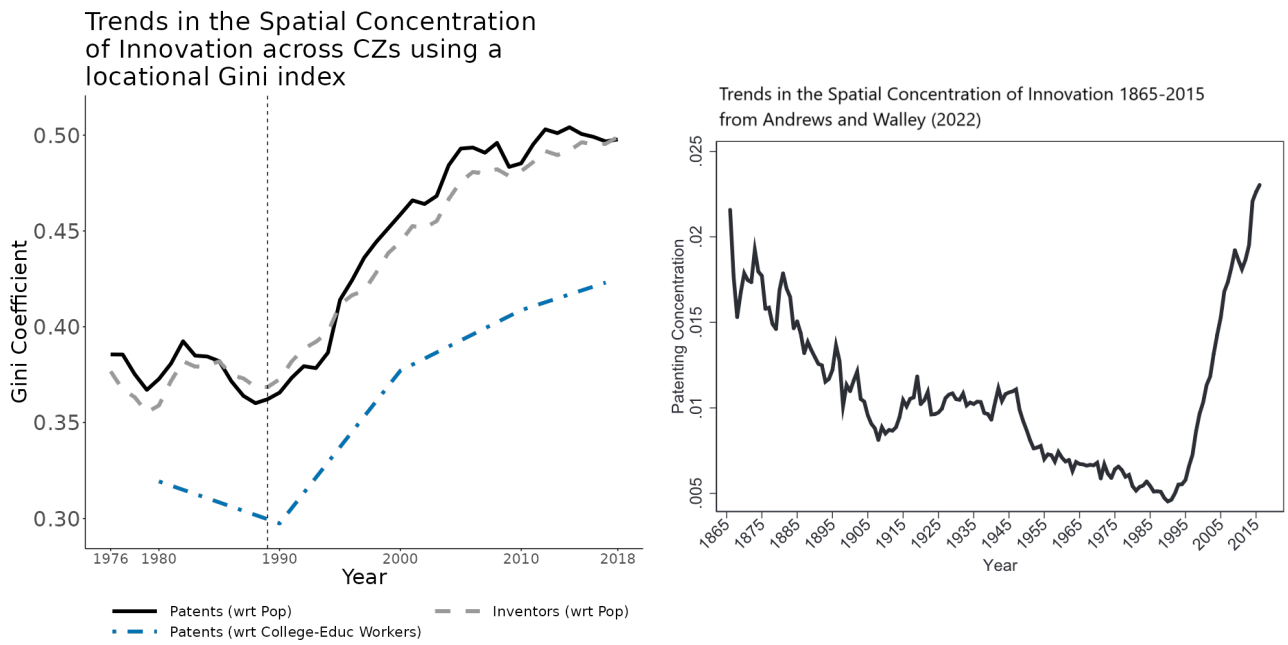


Figure 1: Trends in the spatial concentration of innovation across CZs. The left panel plots trends in the spatial concentration of innovation from 1976 to 2018, measured with respect to population and college educated workers using a locational Gini index (Krugman, 1991) and USPV data. For each CZ, I calculate annual shares of U.S. patents, inventors, population, and college-educated workers. Locational Gini curves are constructed for each year by ranking CZs based on the ratio of their patent or inventor share to their population or college-educated worker share, and plotting the cumulative sum of patent or inventor shares against the corresponding population or college-educated worker shares. The locational Gini index for each year is calculated as twice the area between the 45-degree line and the respective locational Gini curve. The right panel shows trends in the spatial concentration of innovation from 1865-2015 documented by Andrews and Whalley (2021) using a dartboard innovation intensity concentration index, a normalized measure of the sum of deviations of CZ patent shares from population shares.

geographical sorting of workers by skill from 1980-2000[3]. In Figure 16 in Appendix B.1, I demonstrate the robustness of these trends to: (i) excluding top patenting CZs such as San Jose, San Francisco, Newark, and Los Angeles, and; (ii) employing alternative measures of spatial concentration, including the coefficient of variation, Herfindahl index, simplified Ellison and Glaeser (1997) index, and annual share of patents produced by the top 10 CZs. The right panel presents trends in the spatial concentration of innovation from 1865 to 2015, as documented by Andrews and Whalley (2021)[4] using an alternative patent dataset and a dartboard innovation intensity concentration index: a normalized measure of the sum of deviations of CZ patent shares from population shares building on the Ellison-Glaeser index. This figure also reveals a sharp rise in the spatial concentration of innovation beginning in 1990.

The locational Gini curve also provides a natural, annual measure of patenting activity for each CZ that allows for meaningful comparisons over time: the patent-to-population share ratio. This measure reflects patents per capita while normalizing the total number of patents and population in each year, effectively capturing each CZ's contribution to the overall locational Gini coefficient in that year. Compared to existing

[3]I compare this trend against the geographical sorting of patents in Fact 10

[4]Note that this paper does not analyze where and why innovation became more spatially concentrated starting in 1990.

measures in the literature, changes in the patent-to-population share ratio over time reveal whether patents are increasingly concentrated in a given CZ, independent of scale and after controlling for changes in local and national populations[5]. Figure 2 depicts the geography of changes in the patent-to-population share ratio between 5-year averages around 1990 and 2015 while Table 1 in Appendix B.1 lists the top 15 CZs with the largest increase in this period. In addition to well-known superstar cities such as San Jose, San Franciso, San Diego, Seattle, and Boston, regions like Portland, Boise, Wayne, Provo, and Fort Collins also emerged as some of the most innovative areas in the United States between 1990 and 2015. Figure 3 shows how the geography of these changes relate to the rise in the Gini coefficient since 1990. The left panel displays the patent-to-population share ratio across regions in 1990, while the right panel shows that regions with a higher ratio in 1990 generally maintained an even higher ratio in 2015. Conversely, regions with a lower ratio in 1990 experienced an even lower ratio by 2015.

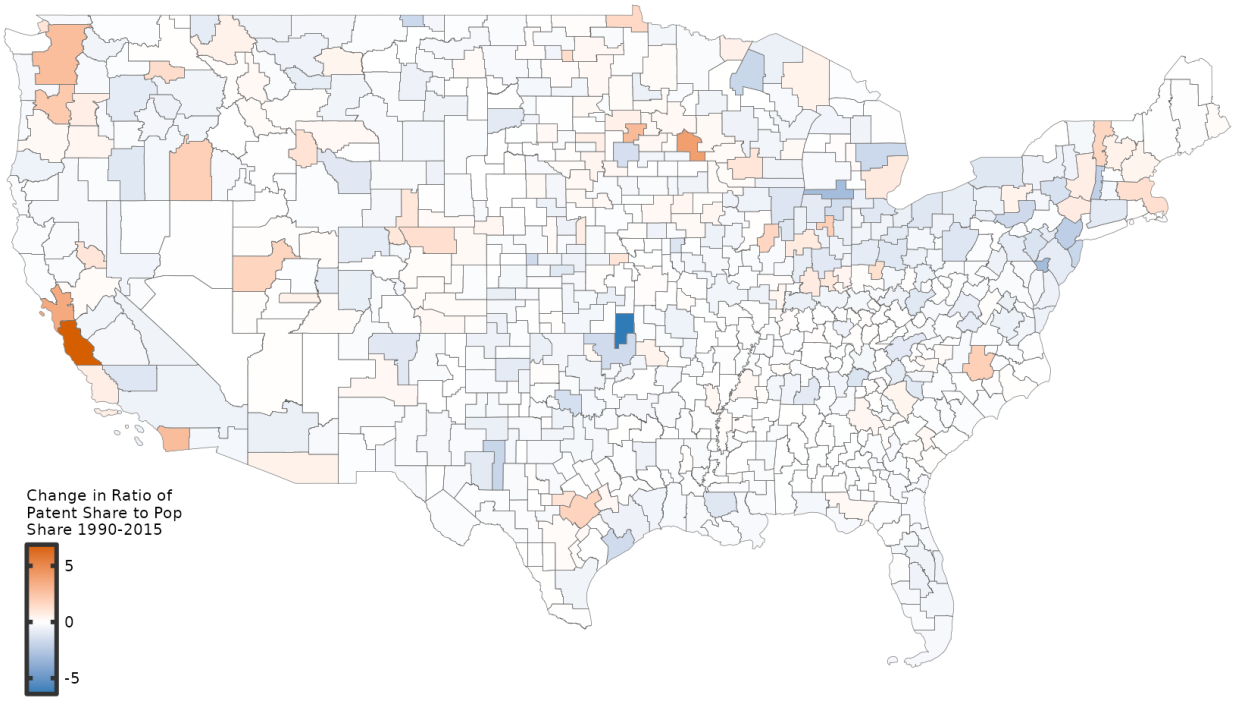


Figure 2: Changes in the geography of innovation intensity from 1990 to 2015, measured using changes in the patent-to-population share ratio on a pseudo-log scale.

[5]Raw patent counts, commonly used in other papers, suffer from several limitations. First, they inherently favor regions with larger populations, leading to biased rankings that can change arbitrarily when regions are grouped differently. Second, intertemporal comparisons are less informative because increases in raw patent counts can result from aggregate growth in patenting, changes in a CZ's share of annual patents, or regional population growth. R&D expenditures, an alternative measure of innovation, are typically available only at the state level.

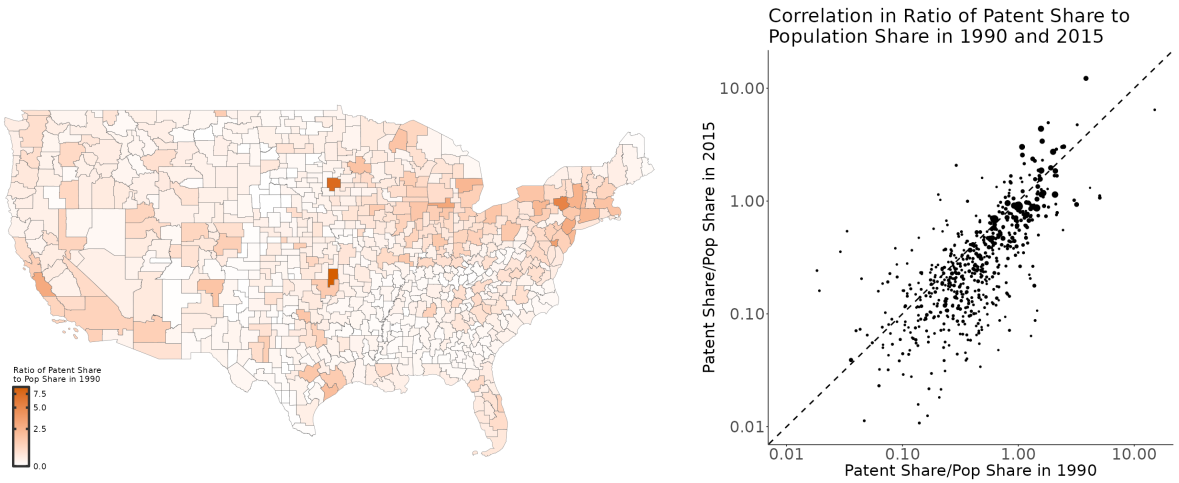


Figure 3: The geography of the patent-to-population share ratio in 1990 (left) and its correlation with the same ratio in 2015 (right)

Using this measure, Figure 4 shows that innovation became increasingly concentrated in high-skill CZs between 1990 and 2018. The left graph plots the correlation between the log college ratio in 1990 and the percentage change in the patent-to-population share ratio from 1990 to 2015 across CZs. Specifically, a 1% increase in the college ratio in 1990 is associated with a 0.8% greater increase in the patent-to-population share ratio over this period. To account for regions with zero patents and to exploit the annual frequency of the dataset, I estimate annual elasticities α_t of patents per capita[6] with respect to the 1990 college ratio using the following Poisson Pseudo Maximum Likelihood (PPML) specification:

$$\text{Patents per capita}_{r,t} = \exp\left(\alpha_t \cdot \text{Log 1990 College Ratio}_r \times \text{Year}_t + \gamma_t + \epsilon_{r,t}\right). \tag{1}$$

where r represents regions and t denotes years. The estimated annual aggregate elasticities α_t serve an important role for the remainder of my empirical analysis. The right graph of Figure 4 shows that this elasticity remained fairly constant at around 1 from 1976 to 1990 but rose sharply thereafter. This trend reflects the gradual geographical sorting of patents per capita across CZs, driven by differences in their initial college ratios in 1990. Importantly, the geographical sorting of patents per capita by initial skill ratio is not mechanically driven by greater increases in the college ratio of initially high-skill CZs: in Figure 15 in Fact 10, I document that worker sorting by skill in the US predominantly occurred in the 1980s and did not continue after 1990. Additionally, Figure 17 in Appendix B.1 plots trends in the elasticity of CZ patents per capita with respect to the 1990 population and population density, respectively. The absence of a clear break in these trends after 1990 suggests that the skill mix of workers in 1990 is a more important margin for the geographical sorting of patents per capita over time than either population size or density.

[6]This is equivalent to the patent-to-population share ratio with year fixed effects.

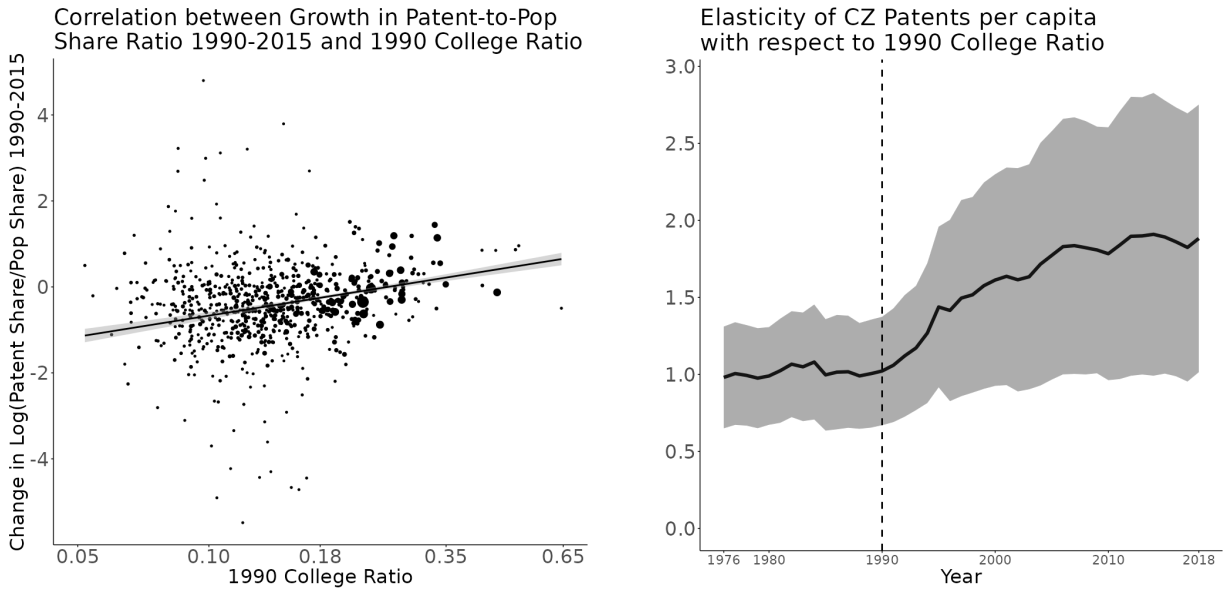


Figure 4: The geographical sorting of patents per capita into high-skill cities from 1990. The left graph shows how percentage changes in the patent-to-population share ratio from 1990 to 2018 relate to the initial college ratio across CZs. Specifically, it plots changes in log patent-to-population share ratio between five-year averages around 1990 and 2015 against the 1990 college ratio on a log scale, with the size of each CZ's dot representing its population. The right graph plots trends in the annual elasticity of CZ patents per capita with respect to the 1990 college ratio. The confidence intervals in both graphs reflect heteroskedastic-robust standard errors.

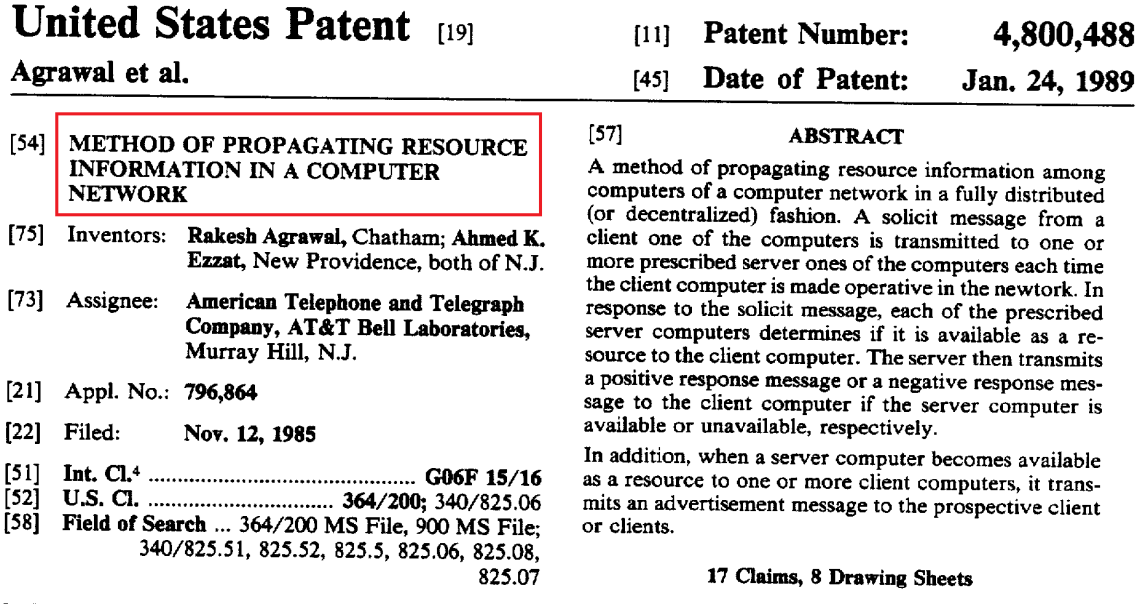
Fact 2: Information and communication technologies rose rapidly after 1990 (the ICT shock)

Why did innovation become more spatially concentrated in high-skill cities after 1990 but not before? I find that the sudden rise of information and communication technologies (the ICT shock) around 1990 accounts for much of this trend. The ICT shock can be characterized by two key aspects. First, there was a dramatic compositional shift in innovation towards the ICT sector. Kelly et al. (2021) identify eight breakthrough patents[7] related to computer networks between 1985 and 1995, with the first four produced during 1985-1990. These influential patents served as a catalyst for the ICT boom in innovation, as evidenced by the similarity in content and the citations made by subsequent ICT patents. Figure 5 highlights the content of the first of these breakthrough patents, which introduced a method for propagating resource information in a computer network.

This first, direct aspect of the ICT shock, i.e. a sharp rise in the annual share of ICT innovation in the ICT sector after 1990, is illustrated in the left panel of Figure 6, which plots trends in the annual number of patents by technology field[8]. While patenting has generally increased over time across all fields, the growth in ICT patents from 1990 to 2018 is particularly pronounced, rising from approximately 8.5% in 1990 to 33% in 2018 (middle panel).

[7]these are Patent Numbers 4,800,488; 4,823,338; 4,827,411; 4,887,204; 5,249,290; 5,341,477; 5,544,322; and 5,586,260

[8]See Appendix A for more details on patent technology fields and subfields.



United States Patent [19]

Agrawal et al.

[11] **Patent Number: 4,800,488**

[45] **Date of Patent: Jan. 24, 1989**

[54] **METHOD OF PROPAGATING RESOURCE INFORMATION IN A COMPUTER NETWORK**

[75] Inventors: **Rakesh Agrawal**, Chatham; **Ahmed K. Ezzat**, New Providence, both of N.J.

[73] Assignee: **American Telephone and Telegraph Company, AT&T Bell Laboratories**, Murray Hill, N.J.

[21] Appl. No.: **796,864**

[22] Filed: **Nov. 12, 1985**

[51] **Int. Cl.**[4] **G06F 15/16**

[52] **U.S. Cl.** **364/200**; 340/825.06

[58] **Field of Search** ... 364/200 MS File, 900 MS File; 340/825.51, 825.52, 825.5, 825.06, 825.08, 825.07

[57] **ABSTRACT**

A method of propagating resource information among computers of a computer network in a fully distributed (or decentralized) fashion. A solicit message from a client one of the computers is transmitted to one or more prescribed server ones of the computers each time the client computer is made operative in the newtork. In response to the solicit message, each of the prescribed server computers determines if it is available as a resource to the client computer. The server then transmits a positive response message or a negative response message to the client computer if the server computer is available or unavailable, respectively.

In addition, when a server computer becomes available as a resource to one or more client computers, it transmits an advertisement message to the prospective client or clients.

17 Claims, 8 Drawing Sheets

Figure 5: Contents of the first of eight breakthrough patents in ICT identified by Kelly et al. (2021)

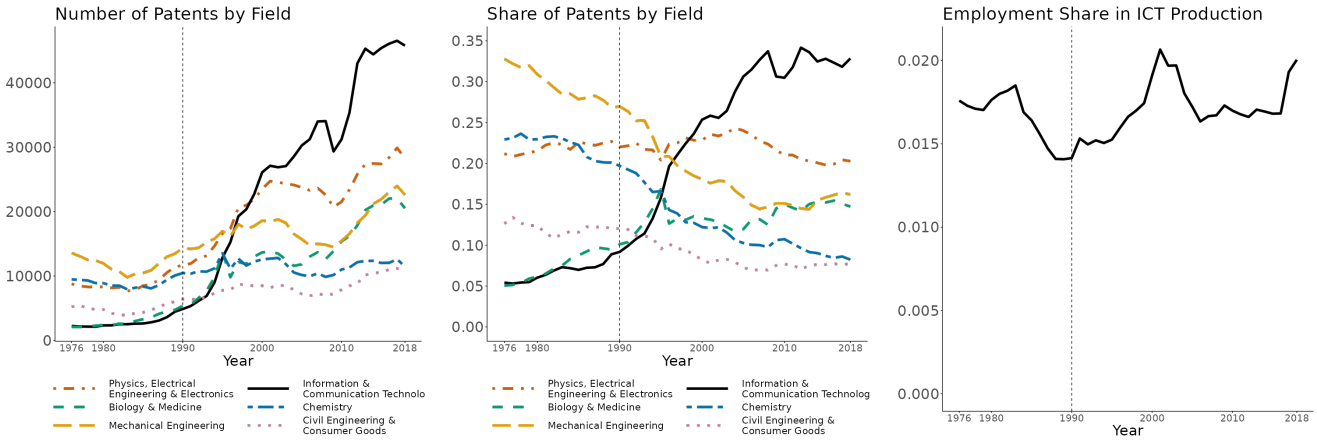


Figure 6: Trends in the number of patents (left) and share of annual patents (middle) by technology field, as well as the national employment share in ICT production (right).

In contrast, trends in employment in ICT production are less clear. Building on Fort et al. (2020), I define ICT production to include the following industries in the Information Sector (NAICS 51): Software Publishers (5112); Telecommunications (517), and; Data Processing, Hosting, and Related Services (518). I categorize these industries as ICT *production*, rather than ICT *innovation*, because establishments that focus on R&D activities are instead classified as Scientific Research and Development Services (5417) and occasionally Corporate, Subsidiary, and Regional Managing Offices (551114) under the North American Industry Classification System (NAICS). The right panel of Figure 6 illustrates trends in the national employment share in ICT production, based on harmonized County Business Patterns data from Eckert et al. (2020). The graph shows that the employment share in ICT production rose from approximately 1.4% in 1990 to 2.0% in the early 2000s, but it declined in the subsequent years as well as during the 1980s.

Second, the ICT shock significantly reduced communication costs across regions. Greenstein (2015) and Jiang (2023) characterize the ICT revolution as the rising availability of high-speed internet, facilitated by the development of the National Science Foundation Network (NSFNET) from 1986 until its full privatization in 1995. The NSFNET, established by the National Science Foundation, was designed to connect supercomputing centers across the US and provide high-speed internet to researchers nationwide. A pivotal moment came in **March 1991**, when the Acceptable Use Policy was modified to allow commercial traffic on the network, granting firms access to high-speed internet for the first time. Although the NSFNET backbone consisted of just 11 nodes in 1991 and 15 in its full version in 1993, many of these nodes connected to regional networks, thereby extending high-speed internet access to universities and firms across most US regions. Appendix B.2 provides a detailed history of the NSFNET.

Fact 3: ICT innovation is more concentrated in high-skill cities compared to other fields

The compositional shift of innovation toward ICT – i.e., the direct aspect of the ICT shock – increased the aggregate spatial concentration of innovation, as ICT patents are more concentrated in high-skill cities compared to patents in other fields. Figure 7 illustrates this disparity: both the Gini coefficient of patents per capita (left graph) and the elasticity of patents per capita with respect to the 1990 college ratio (right graph) are notably higher in ICT (as shown by the solid black lines) compared to other fields before and after 1990.

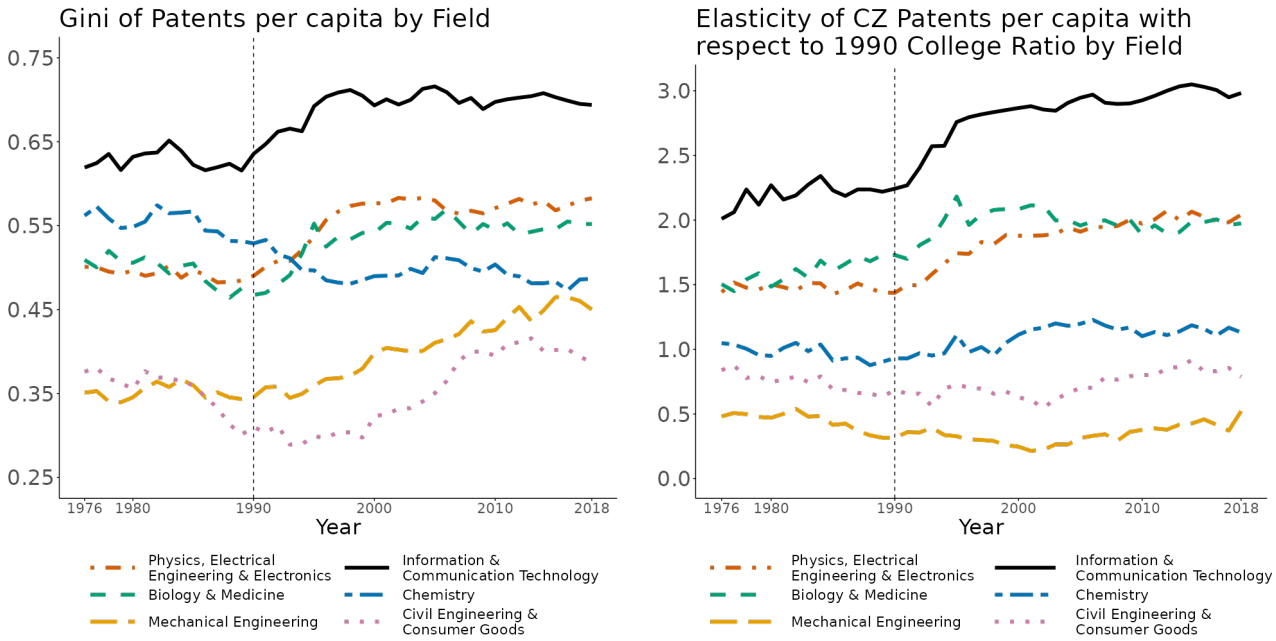


Figure 7: Trends in the locational Gini coefficient (left) and elasticity with respect to 1990 college ratio (right) of patents per capita by technology field.

More formally, I decompose the post-1990 increase in the aggregate annual elasticity, α_t, of CZ patents per capita with respect to the 1990 college ratio into within-field and cross-field components as follows:

$$\alpha_{t^*} - \alpha_{1990} = \sum_{t=1991}^{t=t^*} \Delta\alpha_t = \sum_{t=1991}^{t=t^*} \Bigg[\underbrace{\sum_k \overline{\alpha}_{k,t} \Delta s_{k,t}}_{\substack{\text{changes in field} \\ \text{composition}}} + \underbrace{\sum_k \overline{s}_{k,t} \Delta\alpha_{k,t}}_{\substack{\text{within-field} \\ \text{changes}}} + \underbrace{\Delta\Big(\alpha_t - \sum_k s_{k,t}\alpha_{k,t}\Big)}_{\substack{\text{residual: changes in the} \\ \text{colocation of fields}}} \Bigg] \tag{2}$$

where $\alpha_{k,t}$ is the annual elasticity of CZ patents per capita in field k and year t with respect to the 1990 college ratio, and $\overline{x}_t = \frac{x_t + x_{t-1}}{2}$ and $\Delta x_t = x_t - x_{t-1}$ denote the average and change of any variable x between $t-1$ and t. The first term reflects the impact of changes in the field composition of US patents. This term is positive if patents are increasingly produced in fields that are more spatially concentrated in high-skill cities. The second term captures the role of changes in the spatial concentration of patents in high-skill cities within fields. The third term accounts for changes in the colocation of fields in high-skill cities, measured by differences in the overall elasticity of patents per capita against the 1990 college ratio from a weighted mean of the field-specific elasticities. The left panel of Figure 8 plots trends in this decomposition, while the right panel further decomposes the cross-field component into contributions from individual fields. These graphs show that 60% of the increase in the overall elasticity is attributed to cross-field changes – 53% of the increase in the overall elasticity is explained by the rising share of ICT patents and 7% from the rising share of Biology patents. In Appendix B.3.1, I decompose the rise in the overall Gini coefficient of patents per capita from 1990 to 2018 and find that 52% of the increase is similarly driven by the rising annual share of ICT patents.

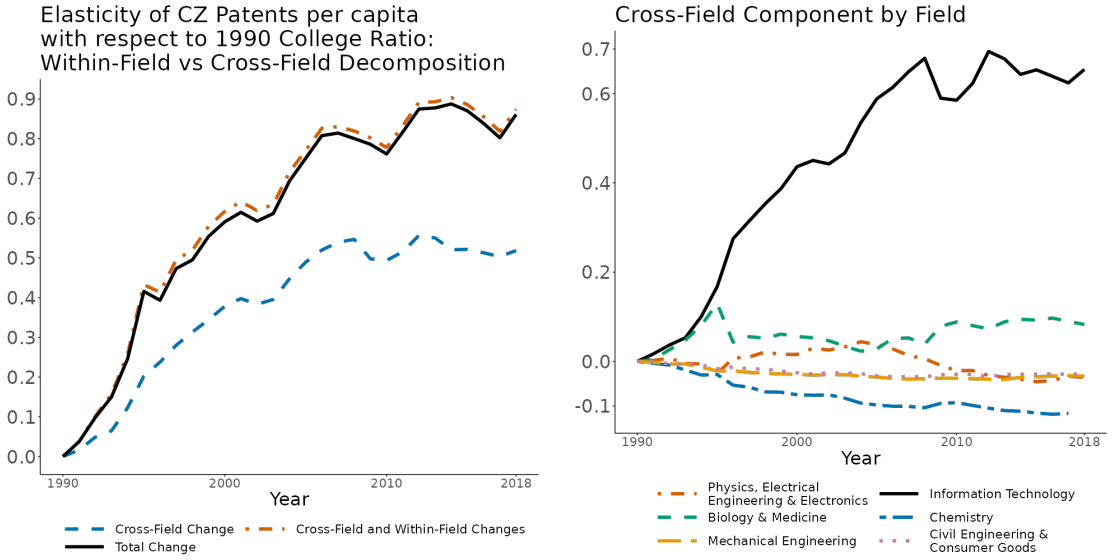


Figure 8: Trends in the decomposition of the aggregate elasticity of CZ patents per capita with respect to the 1990 college ratio. The left panel decomposes the aggregate elasticity into within versus cross field components, while the right panel further breaks down the cross-field component into contributions from individual fields.

Fact 4: ICT innovation is *colocated* with ICT production, both of which became more concentrated in high-skill cities after 1990

An underlying mechanism that explains the greater concentration of ICT innovation in high-skill cities relative to other fields is its *colocation* with ICT production. Specifications (1) and (3) in the left panel of Figure 9 provides evidence of this mechanism from the following regression:

$$\text{ICT Patents per capita}_{r,t} = \beta \cdot \text{ICT Employment Share}_{r,t} \times 1(\text{Year} \geq 1990) + \gamma_t + \varepsilon_{r,t}, \tag{3}$$

where r represents CZs, t represents years, γ_t are year fixed effects.

Dependent Variable:	ICT patents per capita		Non-ICT patents per capita	
Model:	(1)	(2)	(3)	(4)
Log ICT Emp Share	0.9662***	-0.2483	0.2707***	0.0856
× Before 1990	(0.1266)	(0.1828)	(0.1050)	(0.1025)
Log ICT Emp Share	1.988***	0.4262***	0.6292***	0.2408***
× After 1990	(0.3045)	(0.1604)	(0.1438)	(0.0927)
Fixed-effects				
Year	Yes	Yes	Yes	Yes
CZ		Yes		Yes
Fit statistics				
Observations	29,602	25,748	29,602	29,397
Squared Correlation	0.00094	0.91117	0.02510	0.78363
Pseudo R^2	-839,527.3	0.41004	-615,422.7	0.19299
BIC	0.3	0.3	0.3	0.4

Clustered (CZ) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

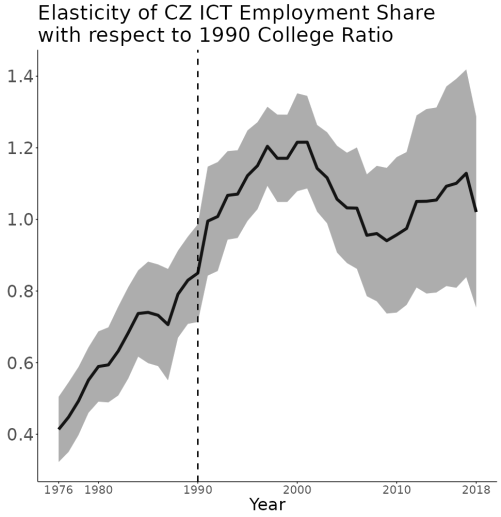


Figure 9: Colocation of ICT patents with employment share in ICT production, which is increasingly concentrated in high-skill cities. The left table shows the correlation between ICT/non-ICT patents per capita and the employment share in ICT production with year and CZ fixed effects. Each observation is at the CZ-year level, weighted by CZ population. The right graph plots trends in the annual elasticity of the CZ employment share in ICT production with respect to the 1990 college ratio.

Before 1990, a 10% increase in the employment share in ICT production is associated with a 9.6% increase in ICT patents per capita and a 2.7% increase in non-ICT patents per capita annually across CZs. The statistically significant difference between these estimates highlights the stronger colocation of ICT production with ICT innovation compared to non-ICT innovation. After 1990, a 10% increase in the employment share in ICT production is associated with a 19.9% increase in ICT patents per capita and a 6.3% increase in non-ICT patents per capita annually across CZs. That ICT patents per capita is much more strongly correlated with ICT production relative to non-ICT patents per capita suggests an even stronger colocation between the ICT production and innovation after 1990[9].

Colocation explains why ICT innovation is more concentrated in high-skill cities compared to non-ICT innovation, as the ICT service sector is predominantly concentrated in these regions. The right panel of Figure 9 plots trends in the annual elasticity of the CZ employment share in ICT production with

[9]Results from the estimation of these relationships at the firm-CZ level are awaiting disclosure from the US Census Bureau.

respect to the 1990 college ratio, captured by the estimated γ_t coefficients from the following regression via PPML:

$$\text{ICT Emp Share}_{r,t} = \exp\left(\gamma_t \cdot \text{Log 1990 College Ratio}_r \times \text{Year}_t + \gamma_t + \epsilon_{r,t}\right). \quad (4)$$

where r represents CZs, t represents years, and γ_t are year fixed effects. These elasticities are positive even before 1990, with a 10% increase in the 1990 college ratio across CZs associated with a 4 to 8% higher employment share in the ICT production in various years prior to 1990.

Additionally, employment in ICT production became increasingly concentrated in high-skill cities from 1990 to 2000. The right panel of Figure 9 shows that the annual elasticity of the ICT employment share with respect to the 1990 college ratio increased from approximately 0.8 in 1990 to 1.2 in 2000. This rising concentration of ICT production employment is correlated with the increasing concentration of ICT innovation in these cities. Specifications (2) and (4) in the left panel of Figure 9, which incorporate both year and CZ fixed effects in equation 3, indicate that after 1990, a 10% increase in the employment share of the ICT service sector within a given CZ over time is associated with a 4.3% increase in ICT patents per capita and a 2.4% increase in non-ICT patents per capita, after controlling for national trends[10]. In turn, the increasing concentration of ICT innovation in high-skill cities since 1990 accounts for an additional 13% of the overall rise in the patent concentration during this period, as demonstrated in my field decomposition using equation 2.

In summary, these findings extend the results of Fort et al. (2020) by offering an ICT sector-specific analysis of the colocation of innovation and production. Together, they underscore *colocation* as a critical mechanism driving both the greater concentration of ICT innovation in high-skill cities compared to non-ICT innovation and the subsequent increase in the ICT innovation concentration in high-skill cities since 1990.

Fact 5: Non-ICT innovation became more concentrated in high-skill cities after 1990

Facts 3 and 4 focus on the role of the ICT sector in driving the rising spatial concentration of aggregate innovation. An important question that remains from my field decomposition in equation 2 and Figure 8 is why non-ICT innovation also became increasingly concentrated in high-skill cities, accounting for the remaining one-third of the overall rise in innovation concentration in these regions since 1990. To investigate this, I first document trends in the spatial concentration of innovation within each field. Figure 10 replicates the trends in the Gini coefficient and elasticity by field from Figure 7, normalizing the 1990 levels to zero for clearer comparison over time.

The left panel of Figure 10 shows that the locational Gini coefficient increased for all fields except Chemistry between 1990 and 2018. Notably, the spatial concentration of patents in Physics, Electrical Engineering & Electronics, and Biology & Medicine rose sharply during the 1990 to 2000 period. These fields predominantly concentrated in high-skill cities, as evidenced by the right graph: the annual elasticity of CZ patents per

[10]Results from the estimation of these relationships at the firm-CZ-year level are awaiting disclosure from the US Census Bureau.

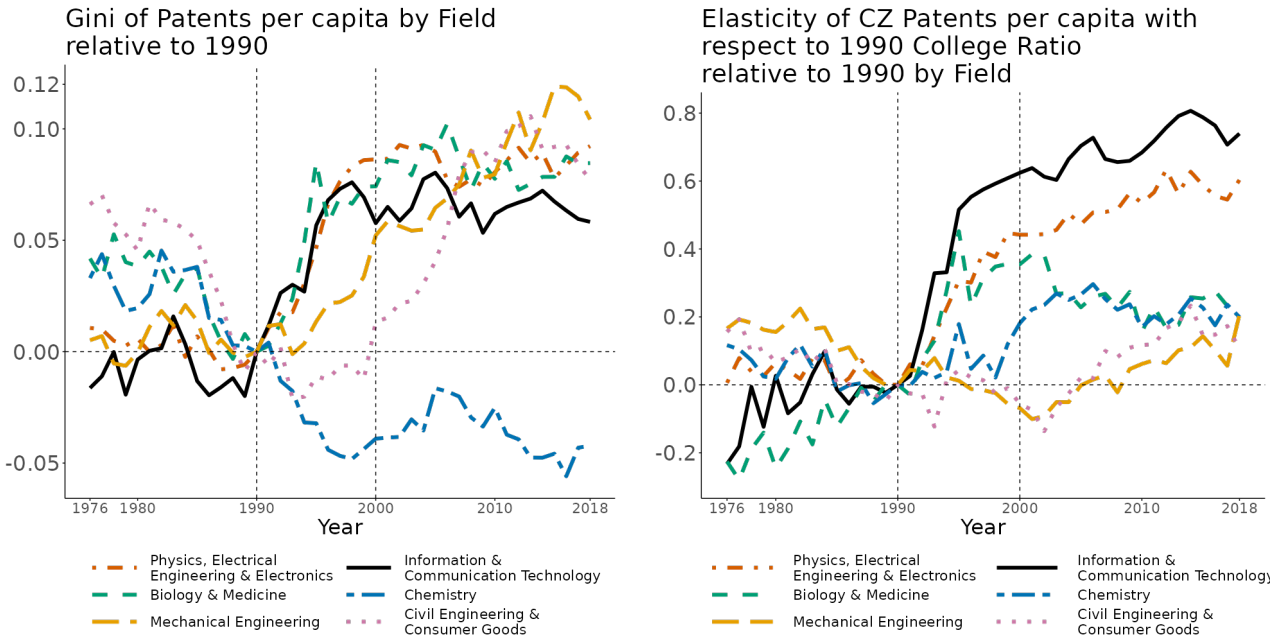


Figure 10: Trends in changes relative to 1990 of the locational Gini coefficient (left) and elasticity with respect to 1990 college ratio (right) of patents per capita by technology field.

capita in these fields with respect to the 1990 college ratio increased significantly over the same period. In contrast, the spatial concentration of patents in Mechanical Engineering rose more gradually from 1990 to 2018, while the spatial concentration of patents in Civil Engineering & Consumer Goods primarily increased between 2000 and 2018. Patents in these fields also became increasingly concentrated in high-skill cities from the early 2000s to 2018[11]. The yellow long-dashed line and the pink dotted line in the right graph indicate that the elasticity of patents per capita in these fields with respect to the 1990 college ratio increased steadily during this period, though to a smaller extent compared to the sharper rises observed for Physics, Electrical Engineering & Electronics, and Biology & Medicine.

Fact 6: Firms initially concentrated in high-skill cities drove non-ICT patent growth after 1990

The rising concentration of non-ICT innovation in high-skill cities since 1990 could be driven by firm entry, compositional changes in patent shares across existing firms, and changes in the geography within firms. Identifying their relative importance is challenging. To make progress, I first decompose the rising spatial

[11]More precisely, I find that patents in Mechanical Engineering became more concentrated in manufacturing hubs from 1990 to 2000 and shifted toward high-skill cities from the early 2000s onwards. Supporting evidence for the rising concentration of Mechanical Engineering patents in manufacturing hubs during this earlier period is available on request.

concentration of non-ICT patents in high-skill cities α_t into within-firm and across-firm components:

$$\alpha_{t^*} - \alpha_{1990} = \sum_{t=1991}^{t=t^*} \Delta\alpha_t = \sum_{t=1991}^{t=t^*} \Bigg[\underbrace{\sum_f \overline{\alpha}_{f,t} \Delta s_{f,t}}_{\substack{\text{changes in firm} \\ \text{composition}}} + \underbrace{\sum_f \overline{s}_{f,t} \Delta\alpha_{f,t}}_{\substack{\text{within-firm} \\ \text{changes}}} + \underbrace{\Delta\Big(\alpha_t - \sum_f s_{f,t}\alpha_{f,t}\Big)}_{\substack{\text{changes in the} \\ \text{colocation of firms}}} \Bigg], \tag{5}$$

where $\alpha_{f,t}$ is the annual elasticity of CZ patents per capita in firm f and year t with respect to the 1990 college ratio[12]. Analogous to the field decomposition in equation 2 above, the first term captures the impact of changes in the composition of patenting firms, the second term reflects changes in the concentration of patents in high-skill cities within firms, and the third term accounts for the residual. The left panel of Figure 11 plots trends from this decomposition, showing that the rising concentration of non-ICT innovation in high-skill cities is almost entirely driven by the first term, compositional changes across firms.

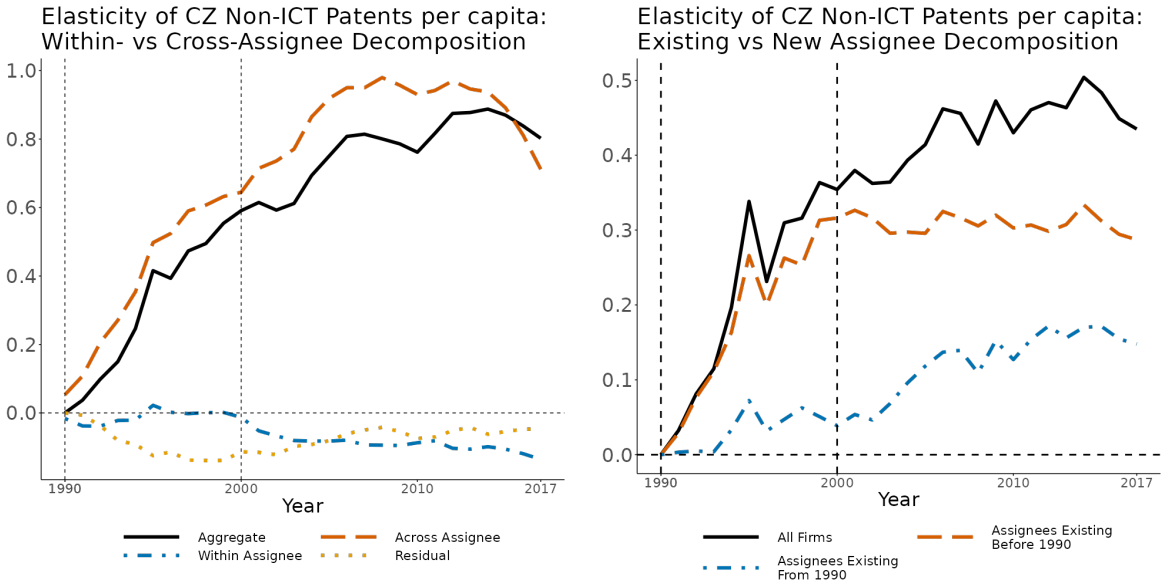


Figure 11: Decomposition of the aggregate elasticity of CZ non-ICT patents per capita with respect to the 1990 college ratio into within versus across firm (left) and into firms existing before 1990 versus new firms (right).

To assess whether these compositional changes are driven by firm entry, I further decompose the rising spatial concentration of non-ICT patents in high-skill cities into contributions from firms that began patenting before 1990 versus those that started afterward. Specifically, I categorize patents into these two mutually exclusive groups and calculate the impact of changes in the concentration of patents within each group, as well as the effect of the rising share of patents in the latter group. The right panel of Figure 11 presents trends from this decomposition, revealing that most of the increase is driven by firms that began patenting before 1990[13].

[12] I acknowledge that some of these firm-level elasticities are estimated with large standard errors, which are omitted from my decomposition. Nonetheless, the decomposition across firms is primarily descriptive. Future drafts will incorporate more econometrically robust methods; however, the central finding – that compositional changes across firms account for nearly all of the rising concentration of non-ICT patents in high-skill CZs since 1990 – is unlikely to change.

[13] I attribute both the rising share of patents from firms that began patenting from 1990 and the rising concentration of

A direct implication of the results from these two decompositions is that firms originally concentrated in high-skill cities produced significantly more patents after 1990. To explore *why* these firms experienced greater patent growth, I conduct a series of firm-level regressions. I first estimate trends in the correlation between firm-level non-ICT patents and their initial patent elasticity (a measure of initial patent concentration in high-skill cities), while controlling for the firm's initial size and spatial scope in 1990. Formally, this trend is represented by δ_t in the following specification:

$$\text{Non-ICT Patents}_{f,t} = \exp\left(\delta_t \cdot 1990 \text{ Patent Elasticity}_f \times \text{Year}_t + \gamma \cdot \text{Controls}_f + \gamma_t + \epsilon_{f,t}\right) \tag{6}$$

where 1990 Patent Elasticity$_f$ is firm f's average annual elasticity of patents per capita with respect to the 1990 college ratio from 1988 to 1990 (capturing its initial patent concentration in high-skill cities) and similar to $\alpha_{f,1990}$ in the firm decomposition (equation 5), γ_t are year fixed effects, γ_f are firm fixed effects, and the set of firm-level time-invariant controls include the firm's average number of annual patents from 1988 to 1990 (a proxy for initial firm size), and the firm's average annual number of CZs of their inventors across all patents from 1988 to 1990 (a proxy for initial firm spatial scope)[14]. The left panel of Figure 12 plots the δ_t coefficients, showing that a one-unit increase in the firm's 1990 patent elasticity with respect to the college ratio is associated with a 20 percentage point greater increase in the number of non-ICT patents in 2000 relative to 1990[15]. This trend suggests that a firm's *initial patent concentration in high-skill cities*, rather than its initial size or spatial scope, explains its subsequent growth in non-ICT patents after 1990.

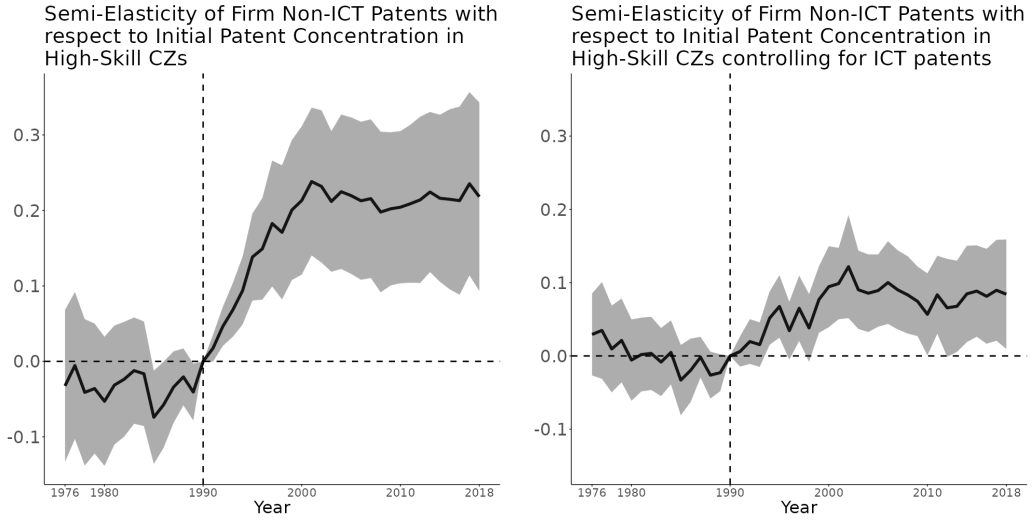


Figure 12: Trends in the semi-elasticity of non-ICT patents with respect to a firm's initial concentration of patents in high-skill cities (1988-1990), and after controlling for the firm's ICT patents in the respective year (right). For each graph, I normalize the estimated semi-elasticity in 1990 to 0.

patents within this group to the contribution of assignee entry (i.e. assignees existing from 1990).

[14]As with the firm decomposition exercise, I acknowledge that the explanatory variable, 1990 Patent Elasticity$_f$, is itself estimated from firm-level regressions and may have large standard errors that are currently ignored. While future drafts will incorporate more econometrically robust specifications and more direct evidence from patent citations, the current results provide circumstantial evidence of the primary mechanisms driving disproportionately larger non-ICT patent growth among firms initially concentrated in high-skill cities. These qualitative findings are unlikely to change.

[15]This trend remains similar with the inclusion of firm fixed effects, and results are available upon request.

Fact 7: Firms initially concentrated in high-skill cities likely benefited more from ICT innovation *spillovers*

Why does a firm's initial patent concentration in high-skill cities in 1990 influence the subsequent trajectory of its non-ICT patents? I provide circumstantial evidence of two potential mechanisms in Facts 7 and 8 respectively. The first mechanism is *spillovers* from ICT to non-ICT innovation. The right panel of Figure 12 plots the δ_t coefficients from equation 6, now incorporating a time-varying control for the firm's trajectory of ICT patents. The trend shows that the increase in firm-level non-ICT patents in the left panel is attenuated by half, implying that approximately 50% of the rise in non-ICT patents is directly associated with corresponding increases in ICT patents within the same firm. While it is possible that other exogenous shocks disproportionately benefited firms in high-skill cities – leading to simultaneous increases in ICT and non-ICT patents after 1990 – identifying alternative shocks that align in both timing and impact with these patterns remains challenging. Instead, the evidence supports the presence of spillovers from ICT to non-ICT innovation, predominantly concentrated in high-skill cities where the ICT boom occurred.

Fact 8: Firms initially concentrated in high-skill cities expanded disproportionately to low-skill regions after 1990 (the *asymmetric scale effect*)

The second mechanism is that firms initially concentrated in high-skill cities disproportionately expanded to new, lower-skill production locations after 1990. I now estimate trends in the correlation between the number of CZs with the firm's establishments and their initial employment elasticity (a measure of initial employment concentration in high-skill cities), while controlling for the firm's initial size and spatial scope in 1990. Formally, this trend is represented by κ_t in the following specification:

$$\text{No of CZs of Plants}_{f,t} = \exp\left(\kappa_t \cdot \text{1990 Emp Elasticity}_f \times \text{Year}_t + \gamma \cdot \text{Controls}_{f,t} + \gamma_t + \gamma_f + \epsilon_{f,t}\right), \quad (7)$$

where 1990 Emp Elasticity$_f$ represents firm f's average annual elasticity of employment per capita with respect to the 1990 college ratio from 1988 to 1990[16], γ_t are year fixed effects, γ_f are firm fixed effects, and the set of controls include the firm's average annual employment from 1988 to 1990 and the firm's average annual number of CZs containing its plants from 1988 to 1990.

The left panel of Figure 13 displays this trend and shows that a one-unit increase in the firm's initial elasticity of employment per capita with respect to the 1990 college ratio is associated with a 4 percentage point greater increase in the number of CZs containing the firm's plants in 2000 relative to 1990. This effect is not an artifact of the regression specification: the middle panel shows trends when the dependent variable in equation (7) is replaced with the firm's total employment, where there are no significant increases after 1990.

[16] I acknowledge that the firm's employment elasticity itself is estimated from a regression of patents within the firm across CZs and may contain large standard errors. Alternative specifications that address this concern will be available in a subsequent draft.

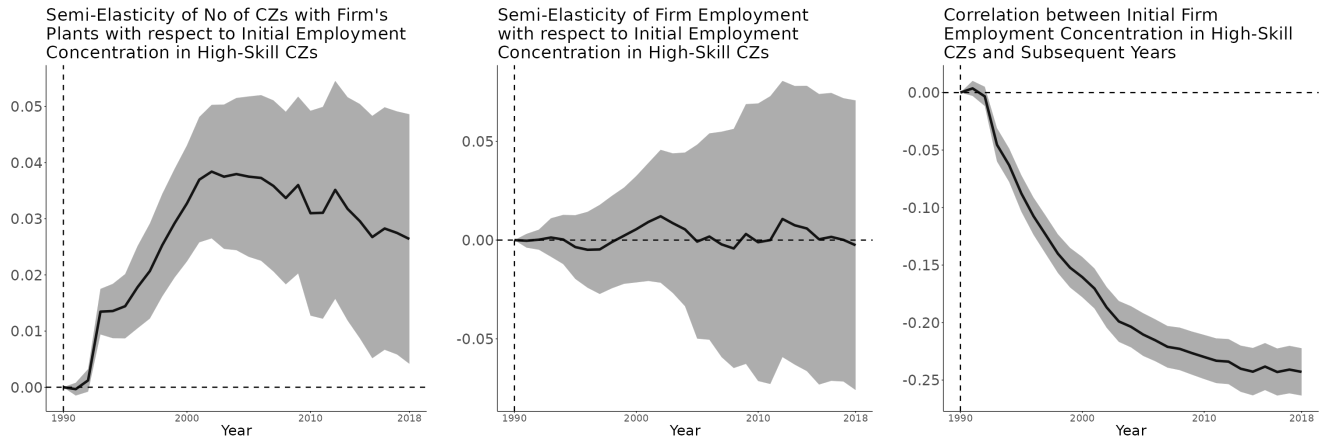


Figure 13: Trends in the semi-elasticity of the number of CZs containing the firm's establishments (left) and the firm's total employment (middle) with respect to the firm's concentration of employment in high-skill cities from 1988 to 1990, along with trends in the correlation between the firm's annual employment elasticity and its initial concentration of employment in high-skill cities (right). In each of these trends, I normalize the 1990 value to 0.

To investigate where firms initially concentrated in high-skill cities expanded to, I examine the correlation between a firm's annual employment elasticity against its initial employment elasticity in 1990, captured by υ_t in the following specification:

$$\text{Emp Elasticity}_{f,t} = \upsilon_t \cdot 1990 \text{ Emp Elasticity}_f \times \text{Year}_t + \gamma \cdot \text{Controls}_f + \gamma_t + \epsilon_{f,t} \tag{8}$$

where $\text{Emp Elasticity}_{f,t}$ is firm f's annual elasticity of employment per capita with respect to the 1990 college ratio and the controls are identical to the ones in equation 7.

The right panel of Figure 13 displays this trend, showing that firms initially concentrated in high-skill cities disproportionately expanded to lower-skill locations. In particular, a one-unit increase in the firm's initial elasticity of employment per capita with respect to the 1990 college ratio is associated with a 0.2-unit greater decrease in the firm's elasticity of employment per capita with respect to the 1990 college ratio by 2000 relative to 1990. This trend suggests that while firms maintained their presence in high-skill cities, they increasingly shifted employment and production into lower-skill regions after 1990, just as communication costs fell with the rising availability of high-speed internet from the ICT shock (Fact 2). Viewed through the lens of an Eaton-Kortum model, as formalized in the next section, this shift was likely motivated by cost advantages in these lower-skill areas. By reducing production costs, these firms enhanced the profitability of new ideas, which in turn contributed to the observed rise in non-ICT patents after 1990.

Fact 9: High- and low-skill workers migrated to initially high-skill cities from 1990, keeping relative regional college ratios stable

My analysis thus far has focused on innovation intensity, as measured by the patent-to-population share ratio. Since 1990, workers have been migrating to high-skill cities, amplifying the effects of the ICT shock on the spatial concentration of innovation. The left panel of Figure 14 displays the correlation between population growth from 1990 to 2015 – measured by the change in log points – against the initial college ratio in 1990. The positive slope or elasticity tells us that initially high-skill cities experienced greater population growth from 1990 to 2015. The right panel shows how the slope of this correlation plot varies as the end year on the vertical axis changes from 1990 to 2018, indicating that there has been a rising concentration of population in high-skill cities.

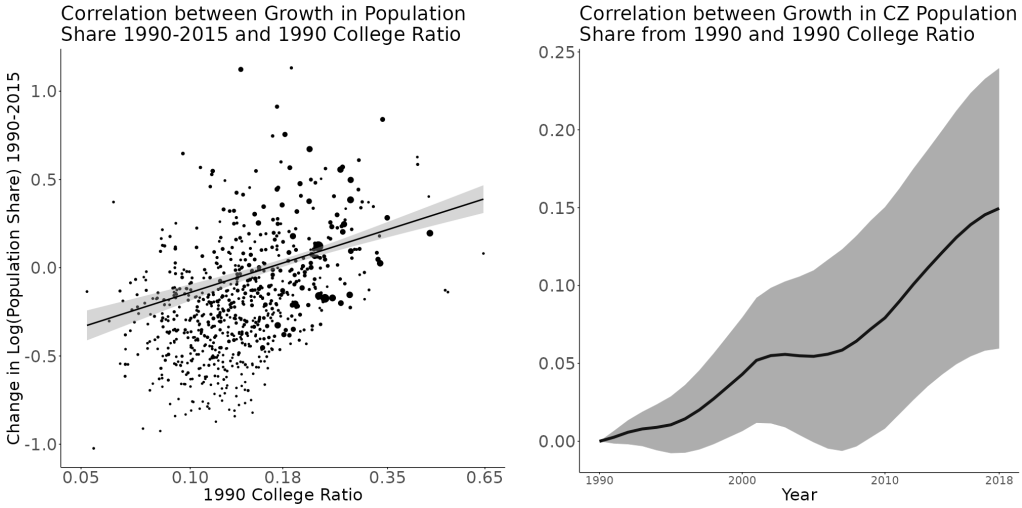


Figure 14: How percentage changes in CZ population share from 1990 to 2015 relate to the initial college ratio. The left panel displays changes in the log population share between five-year averages around 1990 and 2015 against the 1990 college ratio. Each CZ is weighted by its initial population. The right panel shows how the slope of this correlation plot varies as the end year on the vertical axis changes from 1990 to 2018, without taking five-year averages. The confidence intervals capture heteroskedastic-robust standard errors.

Despite the rising concentration of population in high-skill cities since 1990, Figure 15 shows that the geographical sorting of workers by skill to these cities *did not occur after 1990*. The left panel shows correlations between the college ratio rank in 1980 and 1990 and shows that the identity of high-skill cities did not change much during this period. The right panel displays the trend in the correlation between the percentage increase in college ratio from 1980 – measured by log points – and the log college ratio in 1980. The point estimate of 0.09 in 2000 indicates that a 10% increase in the 1980 college ratio across CZs is associated with a 0.9% greater increase in the college ratio from 1980 to 2000. This exactly corresponds to Moretti (2013) and Diamond (2016). I extend their analysis by estimating this correlation for 1990, 2010, and 2018, thereby capturing the *dynamics* of worker sorting over time. The absence of an increase in the correlation estimate in 2000, 2010, or 2018 relative to 1990 indicates that worker sorting primarily occurred in the 1980s and did not continue after 1990. This trend suggests that the rising concentration of patents in high-skill cities since 1990 is not merely a mechanical outcome of worker sorting by skill.

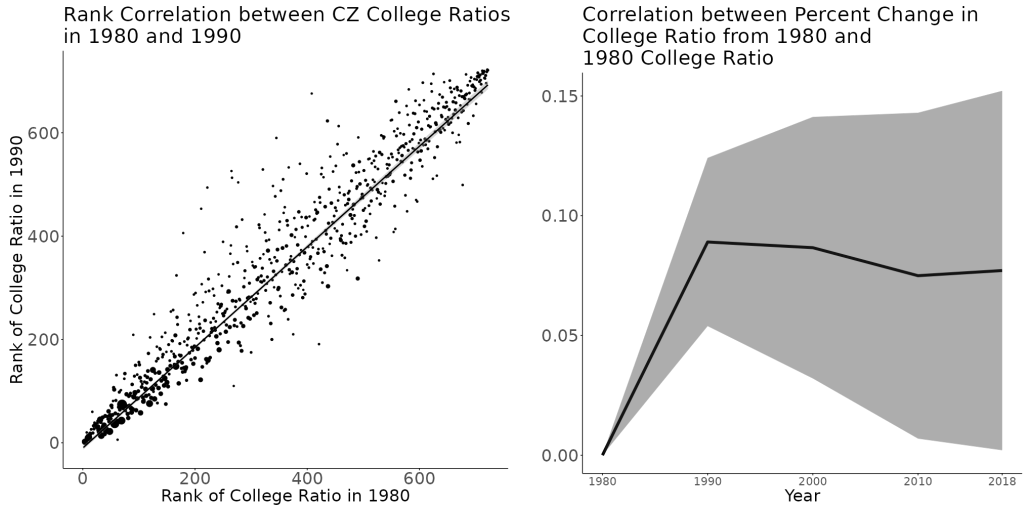


Figure 15: The rank correlation between CZ college ratios in 1980 and 1990 (left) and trends in the correlation between the percentage change in the CZ college ratio from 1980 and the log 1980 college ratio, where each CZ is weighted by its population (right). The confidence intervals capture heteroskedastic-robust standard errors.

3 A Model of Spatial Growth with Endogenous and Directed Innovation, Technology Diffusion, and Dynamic Worker Mobility

My empirical findings suggest that the ICT shock from 1990 (Fact 2) accounts for most of the rising spatial concentration of innovation in high-skill cities after 1990 (Fact 1) through three distinct mechanisms. The first mechanism is a direct effect: a compositional shift of innovation towards ICT, which is ***colocated*** with ICT production (Fact 4) and more concentrated in high-skill cities relative to innovation in other fields (Fact 3). In addition, non-ICT innovation became more concentrated in high-skill cities after 1990 (Fact 5), likely driven by two indirect effects of the ICT shock. Specifically, firms initially concentrated in high-skill cities produced more non-ICT patents from 1990 (Fact 6), likely due to: (i) ***spillovers*** from ICT to non-ICT innovation (Fact 7), and (ii) geographic expansion into lower-cost regions, facilitated by ICT-induced reductions in communication costs (the ***asymmetric scale effect***) (Fact 8). Worker ***migration*** to high-skill cities during this period (Fact 9) amplified the effect of these three mechanisms.

Simultaneously explaining these nine facts – which outline the key mechanisms driving the rising concentration of innovation in high-skill cities – and assessing their aggregate consequences is a significant challenge that demands a highly flexible framework. To address this, I develop a general model of spatial growth that integrates endogenous innovation and technology diffusion at the level of *individual ideas* – the fundamental unit of the Eaton-Kortum structure – along with worker mobility across regions, sectors, and between production and research. Specifically, I first leverage tools from Eaton and Kortum (2024) to introduce imperfect competition and, consequently, endogenous innovation in Lind and Ramondo (2024)'s quantitative trade model with idea diffusion. I then incorporate dynamic worker mobility with frictions by using an extended version of Caliendo et al. (2019). These model components allow me to make two methodological contributions. First, I *introduce directed innovation*[17] *in a quantitative spatial model*, generating an endogenous and gradual response of spatial innovation patterns following the ICT shock through the three mechanisms outlined in my empirical findings. Second, I provide an *analytical decomposition of the welfare effects of any shock to economic fundamentals into its transitory and long-run growth components*. In deriving this decomposition in the next section, I illustrate how the evolution of the geography of innovation from the ICT shock impacts aggregate growth and welfare.

Concretely, in my model the trade shares and production worker wages are fully endogenous and generated from the dynamics and microfoundations of innovation and technology diffusion, while the incentives to innovate depend on equilibrium trade and technology diffusion. My model features two sectors: ICT and non-ICT. In each region and sector, innovation workers discover new ideas through a Poisson process, where the arrival rate is influenced by various factors. Once an idea is discovered in a particular region, its diffusion to all regions is governed by independent Poisson processes with varying bilateral diffusion speeds. I capture the ICT shock through two exogenous components: (i) an ***increase in the national productivity of ICT innovation***, reflected in the Poisson arrival rate of new ideas; and (ii) a ***decrease in the national***

[17] By directed innovation I refer to the concept of directed technological change introduced by Acemoglu (2002). I extend his work by introducing the *spatial* direction of technological change.

component of bilateral diffusion speeds, representing reduced communication costs. With additional assumptions, my microfounded structure of innovation and technology diffusion generates trade, technology adoption, and idea market shares. These shares underpin two key mechanisms documented in my empirical findings. First, the technology adoption shares determine the degree of ***colocation*** between innovation and production. Second, the idea market shares illustrate the ***asymmetric scale effect***, where a uniform increase in bilateral diffusion speeds disproportionately enhances the market access of ideas discovered in high-skill cities.

To enable these two mechanisms, along with spillovers from ICT to non-ICT innovation, to shape the geography of innovation, I introduce *directed innovation* through imperfect competition – generating profits from production and, consequently, innovation – and worker mobility. In each region and sector, a unit continuum of immobile firms employs innovation workers and owns their ideas. Each firm is thus a collection of ideas. These firms engage in Bertrand competition, where the lowest-cost producer of each good captures its entire market by charging the highest markup that deters competitors from entering, thereby earning profits. The expected value of each idea is determined by the trajectory of its market share, multiplied by the profits from producing the corresponding good. Firms compensate innovation workers with wages equal to the expected value of each idea, multiplied by the Poisson arrival rate of ideas in that region-sector, while production worker wages are determined by the market clearing condition. Given these wages in innovation and production, workers make dynamic decisions about moving between production and innovation and relocating across regions and sectors. The equilibrium worker mobility shares then govern the spatial direction of innovation along the transition path. For my application, I assume that the Poisson arrival rate of new ideas depends on the benefits of ***colocation between innovation and production***, agglomeration economies including ***spillovers from ICT to non-ICT innovation***, and the number of innovation workers in that region and sector (capturing the roles of the ***asymmetric scale effect*** and ***worker migration***). Consequently, these endogenous mechanisms influence innovation worker wages and hence worker mobility patterns, shaping the geography of innovation following the ICT shock, in line with my empirical findings. Finally, the steady state distribution of workers – and the corresponding spatial distribution of innovation rates – ultimately determine the aggregate growth of the economy along the balanced growth path.

Formally, my model consists N regions, denoted by r,o,d, which correspond to the locations of innovation or research (r), production (o for origin), and consumption (d for destination). The two sectors – ICT and non-ICT – are denoted by k,s, while the two types of economic activities are goods production (G) and research (R). Time is continuous, with t^* representing the period when ideas are produced from research and t the period when goods are produced and consumed. The equilibrium objects and central mechanisms of the model are presented as lemmas, propositions, and corollaries, with proofs in Appendix C.

3.1 Microfounded Innovation and Technology Diffusion

In each region and sector, there is a unit continuum of immobile firms. At each time t^*, these firms hire local innovation workers to conduct research, discovering ideas according to a stochastic process at rate $\lambda^k_{r,t^*}dt^*$[18]. To capture the ICT shock and the mechanisms shaping the geography of US innovation, I assume that λ^k_{r,t^*} depends on the region's college ratio, sector-specific national research productivity (capturing a ***one-time increase in ICT research productivity nationwide, the first component of the ICT shock***), the benefits of colocation with production, the number of inventors, and agglomeration economies in innovation (including spillovers from ICT to non-ICT). Note, however, that the remainder of my model can be derived for any arbitrary function of λ^k_{r,t^*}.

Each idea pertains to the production of a specific good ν on the unit interval [0,1] and has stochastic quality determined on discovery, such that the aggregate number of ideas discovered in region r with quality above q by time t is drawn from a Poisson distribution with mean parameter[19]:

$$\lambda^Q_{r,t}(q) = q^{-\theta}\int_{-\infty}^{t}\lambda_{r,t^*}dt^*. \tag{9}$$

After an idea is discovered in region r at time t^*, applications of the idea arrive via independent stochastic processes in every region[20]. Each application has stochastic applicability determined on arrival, such that the aggregate number of applications of this idea in region o at time t with applicability above a is drawn from a separate Poisson distribution with mean parameter[21]:

$$\lambda^A_{ro,t^*t}(a) = a^{-\sigma}\cdot\Omega_{ro,t^*}(t-t^*)\cdot\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}}. \tag{10}$$

From this distribution, the expected number of applications of the idea above a to ever arrive in region o is $\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}}a^{-\sigma}$. $\Omega_{ro,t^*}(t-t^*)$[22] is the share of applications that have arrived in region o by time t and represents any function of the idea's age $t-t^*$ with $\Omega:[0,\infty)\to[0,1]$ and $\lim_{t\to t^*}\Omega_{ro,t^*}(t-t^*)=0$, $\lim_{t\to\infty}\Omega_{ro,t^*}(t-t^*)=1$. Intuitively, $\frac{d\Omega_{ro,t^*}(t-t^*)}{dt}$ captures the density or fertility of the idea in creating new applications. I assume that this fertility declines exponentially with the age of the idea, formalized as follows:

[18] λ^k_{r,t^*} may be thought of as the output of the research production function

[19] Note that I omit the sector superscript k in this equation and what follows since the microfounded structure of innovation and technology diffusion is independent across sectors.

[20] This arrival of applications captures technology diffusion at the level of individual ideas.

[21] I provide microfoundations for the Poisson distributions of idea quality and applicability in Appendix C.1.

[22] Notice that $\Omega_{ro,t^*}(t-t^*)$ could be parameterized as a function of trade and migration flows at time t, mimicking Buera and Oberfield (2020) and Cai et al. (2022). Because innovation and technology diffusion are fundamentally connected at the level of individual ideas in my model, for any *arbitrary* function $\Omega_{ro,t^*}(t-t^*)$, the technology diffusion network then influences innovation rates through the expected value of individual ideas, as shown in Section 3.3 below. This impact cannot be obtained when innovation and technology diffusion are modeled as separate processes.

Assumption 1. *Idea diffusion occurs exponentially over the idea's age, such that:*

$$\Omega_{ro,t^*}(t-t^*) = 1 - e^{-\delta_{ro,t^*}(t-t^*)}, \tag{11}$$

with $\Omega_{rr,t^*}(t-t^*) = 1$ *(i.e. by setting* $e^{-\delta_{rr,t^*}(t-t^*)}$ *as* ∞ *for all* $t \geq t^*$*).*

This assumption strikes a balance between two extremes: instantaneous diffusion, where new ideas are disproportionately weighted, and linear diffusion, where older ideas continue to generate numerous new applications over time. In particular, the speed of exponential diffusion is governed by δ_{ro,t^*}. A decrease in the common component δ_{t^*} across all region-pairs over time captures ***reduced communication costs, the second component of the ICT shock***.

3.2 Evolution of Technology Levels, Production, and Trade

I now show how the discovery and diffusion of individual ideas generate aggregate trade and technology adoption flows across all region-pairs.

3.2.1 Production of Individual Goods

In each region o, local firms produce each individual good ν using only labor and the most efficient idea application available to them, where the efficiency of each idea application in producing its corresponding good ν is given by the product of its quality and applicability. Thus, the productivity of producing each individual good ν in region o at time t is given by:

$$z_{o,t}(\nu) = \max_i \{q_i \cdot a_{i,o,t}\} \tag{12}$$

where q_i is the quality of idea i, and $a_{i,o,t}$ is its best applicability in region o at time t.

Given the stochastic microfoundations of innovation and diffusion detailed in equations (9)-(10), the productivity of each individual good $z_{o,t}(\nu)$ is drawn from a multivariate Fréchet distribution. Formally:

Lemma 1. *Given that each idea is discovered at time* t^* *at a unique discovery location, idea applications have quality and applicability drawn from the Poisson distributions in equations* (9)-(10) *respectively, and their efficiency is multiplicative in quality and applicability [equation 12], the* ***joint productivity distribution*** *of individual goods across regions at each time* t *is multivariate Fréchet and given by:*

$$\mathbb{P}\left[Z_{1,t} \leq z_1, \ldots, Z_{N,t} \leq z_N\right] = \exp\left[-\sum_{o=1}^{N} \int_{-\infty}^{t} \left[\sum_{r=1}^{N} \Omega_{ro,t^*}(t-t^*)\ z_o^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} \lambda_{r,t^*} dt^*\right] \tag{13}$$

where $\rho = 1 - \frac{\theta}{\sigma} < 1$*, and the* ***marginal productivity distribution*** *in each region is Fréchet and given by:*

$$\mathbb{P}\left[Z_{o,t} \leq z_o\right] = \exp\left[-T_{o,t} z_o^{-\theta}\right] \tag{14}$$

with shape parameter $\theta > 0$ and scale parameter:

$$T_{o,t} = \sum_{r=1}^{N} T_{ro,t} = \sum_{r=1}^{N} \int_{-\infty}^{t} \underbrace{\Omega_{ro,t^*}(t-t^*)^{1-\rho}}_{technology\ diffusion} \cdot \underbrace{\lambda_{r,t^*}}_{innovation} \, dt^*. \tag{15}$$

Notice from equation (14) that the marginal productivity distribution in each region is Fréchet, like in the canonical Eaton and Kortum (2002) model [henceforth, EK]. As shown below, this distribution facilitates closed form expressions for trade shares, enhancing tractability. The region's mean productivity in producing individual goods is given by the scale parameter, $T_{o,t}$. Unlike EK, this scale parameter is endogenously determined from the *dynamics* of innovation and technology diffusion, as shown by equation (15). Thus, changes in innovation and technology diffusion patterns from the ICT shock – or any other shock to economic fundamentals – impacts trade shares and, consequently, production worker wages and aggregate welfare. Because technology diffusion is modeled as different applications of the same idea, the impact of technology diffusion on trade shares is governed by the relative variability of idea applicabilities against idea quality, $\rho = \frac{\theta}{\sigma}$. This elasticity is captured by the exponent $1-\rho$ in the marginal (equation 14) and joint productivity distributions (equation 13), ensuring that technology diffusion does not lead to large fluctuations in technology levels and, hence, trade shares over time.

3.2.2 Aggregation across Individual Goods and Equilibrium Trade Shares

In each region and sector, a representative final goods producer aggregates across these individual goods ν via a Cobb-Douglas function[23]:

$$Y_{d,t} = \exp \int_0^1 \ln Y_{d,t}(\nu) d\nu. \tag{16}$$

Each individual good ν need not be produced in the same region d. Instead, the final goods producer purchases each good from its cheapest production location, taking into account trade costs. Trade costs are of the standard iceberg type, such that delivering one unit of any good from region o to region d at time t requires shipping $\tau_{od,t} \geq 1$ units of the good, with $\tau_{oo,t} = 1$ for all o, and $\tau_{od,t} \leq \tau_{od',t}\tau_{d'd,t}$ for all o, d, and d'. Thus, the cost of each good ν in region d is the minimum unit cost across all regions:

$$c_{d,t}(\nu) = \min_{o} \left\{ \frac{\tau_{od,t} w_{o,t}}{z_{o,t}(\nu)} \right\} \tag{17}$$

where $w_{o,t}$ is the wage of production workers in region o at time t.

Consequently, the equilibrium trade shares are given by the share of goods in each destination that are purchased from different production locations. Because the productivity of each good is drawn from a

[23]The Cobb-Douglas aggregator implies that every good is essential for the production of the local sectoral final good, so the monopoly price for each individual good is infinite. Thus, under Bertrand competition introduced in the next subsection, the price charged for every good is a markup on the second lowest cost, simplifying the computation for the expected value of an individual idea. With a CES aggregator, the price of each good is the minimum between the second lowest cost and the monopoly markup of the lowest cost, as shown in Bernard et al. (2003).

multivariate Fréchet distribution, as shown in equation (13), these trade shares can be expressed in closed form. Formally:

Lemma 2. *Given that the productivity of each good is drawn from the multivariate Fréchet distribution in equation* (13)*, equilibrium* ***trade shares*** *are given by:*

$$\pi_{od,t} = \sum_{r=1}^{N} \pi_{rod,t} = \sum_{r=1}^{N} \int_{-\infty}^{t} \underbrace{\phi_{rd,t^*t}}_{\substack{\textit{share of goods in } d \textit{ at } t \\ \textit{using ideas from } r \textit{ at } t^* \\ \textbf{\textit{(idea market shares)}}}} \cdot \underbrace{\varphi_{o|rd,t^*t}}_{\substack{\textit{share of goods in } d \\ \textit{produced in } o \textit{ at } t \\ \textit{given ideas from } r \textit{ at } t^* \\ \textbf{\textit{(conditional idea adoption shares)}}}} dt^* \tag{18}$$

with the ***idea market shares*** *given by:*

$$\phi_{rd,t^*t} = \frac{\Phi_{rd,t^*t}^{1-\rho} \lambda_{r,t^*}}{\sum_{r'} \int_{-\infty}^{t} \Phi_{r'd,t't}^{1-\rho} \lambda_{r',t'} dt'}, \tag{19}$$

where I define the ***idea market access*** *term* Φ_{rd,t^*t} *as:*

$$\Phi_{rd,t^*t} \equiv \sum_{o} \Omega_{ro,t^*} (t - t^*) \left(w_{o,t} \tau_{od,t}\right)^{-\frac{\theta}{1-\rho}}, \tag{20}$$

and the ***conditional idea adoption shares*** *given by:*

$$\varphi_{o|rd,t^*t} = \frac{\Omega_{ro,t^*} (t - t^*) \left(w_{o,t} \tau_{od,t}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N} \Omega_{ro',t^*} (t - t^*) \left(w_{o',t} \tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}}. \tag{21}$$

The trade shares in equation (18) illustrate the interconnections between innovation, diffusion, and trade[24]. Unlike standard trade models, this model allows for innovation and production to occur in different locations. Consequently, trade shares are represented as the sum of trilateral shares, $\pi_{rod,t}$, across all innovation regions r. These trilateral trade shares denote the share of goods sold in destination region d that were produced in origin region o using ideas discovered in region r. Because the discovery time of an idea, t^*, may differ from the time of production, trade, and consumption, t, and because ideas are not perfect substitutes, it is necessary to distinguish goods produced from ideas developed at different times. Accordingly, the trilateral trade shares at time t are expressed as an integral over all idea cohorts $t^* \leq t$: $\pi_{rod,t} = \int_{-\infty}^{t} \pi_{rod,t^*t}\, dt^*$. The trilateral share for each idea cohort, π_{rod,t^*t}, exhibits a nested structure because productivity is drawn from a multivariate Fréchet distribution with correlation across idea discovery locations. Specifically, the trilateral share for each idea cohort equals the share of goods ϕ_{rd,t^*t} sold in destination region d at time t using ideas developed in region r at time t^* (idea market shares) multiplied by the conditional probability $\varphi_{o|rd,t^*t}$ that

[24]Note that the trade shares depend only on the probabilistic assumptions of technology (i.e. the Poisson distributions for idea quality and applicability in equations 9 and 10 respectively) and are independent of market structure. This is because each destination always sources goods from their cheapest location, a fundamental result highlighting the flexibility of the Eaton-Kortum world.

region o is the lowest-cost location for producing these goods (conditional idea adoption shares).

3.2.3 Degree of Colocation between Innovation and Production

The idea market and idea adoption shares, in turn, capture the spatial mechanics of innovation documented in my empirical findings. Specifically, by aggregating the conditional idea adoption shares in equation (21) across different destination markets d, we obtain the (unconditional) idea adoption shares:

$$\varphi_{ro,t^*t} = \sum_d \pi_{o|rd,t^*t} = \sum_d \frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}\tau_{od,t}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}}. \tag{22}$$

This expression represents the share of successful ideas discovered in region r at time t^* that were adopted in region o for production of the corresponding goods at time t. A "successful idea" here refers to one that resulted in a good sold in at least one destination market at time t. Using this expression, I formally characterize the degree of *colocation* between innovation and production in the following corollary:

Corollary 1. *For goods sold at time t using ideas discovered at time t^*, the* ***degree of colocation between innovation and production*** *for any region r relative to an alternative production location $o \neq r$ is given by:*

$$\frac{\varphi_{rr,t^*t}}{\varphi_{ro,t^*t}} = \frac{1}{\Omega_{ro,t^*}(t-t^*)} \cdot \left(\frac{w_{r,t}}{w_{o,t}}\right)^{-\frac{\theta}{1-\rho}} \cdot \frac{\sum_d \left(\tau_{rd,t}\right)^{-\frac{\theta}{1-\rho}}}{\sum_d \left(\tau_{od,t}\right)^{-\frac{\theta}{1-\rho}}}. \tag{23}$$

Generally, this ratio captures the relative share of goods produced in the idea's discovery region r versus an alternative production region o, providing a direct measure of how closely innovation and production are colocated. The different components reflect key factors driving colocation, including the share of all ideas discovered in region r at time t^* that have diffused to o (first term), the cost competitiveness between r and o (second term), and the relative accessibility of destination markets d from r and o (third term). In my application to US innovation, this ratio provides two reasons why ICT production became more concentrated in high-skill cities following a one-time increase in ICT research productivity (the first exogenous component of the ICT shock): frictions in technology diffusion (the first term), and the greater accessibility of ICT destination markets from high-skill cities.

3.2.4 The Asymmetric Scale Effect of Rising Technology Diffusion Speeds

Additionally, the idea market shares (equation 19) characterize the ***asymmetric scale effect***, where high-skill cities experience a greater increase in idea market access (equation 20) from a uniform increase in bilateral diffusion speeds. This is formalized in the following corollary, by substituting Assumption 1 in equation 20:

Corollary 2. *If bilateral diffusion speeds are symmetric ($\delta_{rr',t^*} = \delta_{r'r,t^*} \equiv \delta_{rr'}$) and trade costs are identical across all region pairs ($\tau_{rd,t} = \tau_{r'd,t} \equiv \tau_t$), an increase in the bilateral diffusion speed results in a greater*

increase in idea market access for the region with the higher wage, i.e.:

$$\frac{\partial \Phi_{rd,t^*t}}{\partial \delta_{rr',t^*}} - \frac{\partial \Phi_{r'd,t^*t}}{\partial \delta_{rr',t^*}} = \delta_{rr',t^*} e^{-\delta_{rr',t^*}(t-t^*)} \left[\left(w_{r',t} \tau_{r'd,t} \right)^{-\theta} - \left(w_{r,t} \tau_{rd,t} \right)^{-\theta} \right] > 0 \quad if \quad w_{r,t} > w_{r',t}.$$

Intuitively, an increase in the bilateral diffusion speed provides the higher-skill and hence higher-wage region greater access to the lower-cost region for production of its ideas, exploiting cost efficiencies. In contrast, the lower-wage region experiences smaller gains in idea market access, as its already competitive production costs limit the relative advantage of gaining access to a higher-cost region.

Corollaries 1 and 2 illustrate how the underlying components of the trade shares capture the ***colocation*** of ICT innovation and production in high-skill cities, and the ***asymmetric scale effect***. These endogenous mechanisms can only be obtained in a model where innovation and technology diffusion are connected at the level of individual ideas[25]. To allow these mechanisms, along with ***spillovers*** from ICT to non-ICT innovation, to shape the trajectory of US innovation geography following the ICT shock, I *introduce directed innovation in a dynamic spatial model* through imperfect competition and dynamic worker mobility with frictions. Imperfect competition generates profits from individual ideas that shape the incentives of workers, while their mobility determines the evolution of innovation levels in each region and sector.

3.3 Market Structure, Prices, Profits, and Innovation Worker Wages

In each region and sector, firms hire innovation workers and own their ideas. These firms engage in Bertrand competition, where the lowest cost producer of each good claims the entire market for that good, charging the highest markup that deters any competitor from entering.

3.3.1 Prices and the Markup Distribution

Since all goods are essential in the production of the local final good (equation 16), the price of each good is its second lowest cost. Aggregating across all goods yields the price index in each region. Formally:

Lemma 3. *Given Bertrand competition and the Cobb-Douglas aggregation of goods in each region and sector (equation 16, the* ***price index*** *in region d at time t is:*

$$P_{d,t} = \gamma \left[\sum_{r'=1}^{N} \int_{-\infty}^{t} \Phi_{r'd,t^*t}^{1-\rho} \lambda_{r',t^*} dt^* \right]^{-\frac{1}{\theta}}, \tag{24}$$

where γ is the Euler-Mascheroni constant.

[25] When innovation and technology diffusion are modeled at an aggregate level (e.g. Desmet et al., 2018) or as independent processes (e.g. Buera and Oberfield, 2020; Cai et al., 2022), the idea *adoption* shares would exactly coincide with the idea *diffusion* shares, so the degree of colocation between innovation and production does not depend on wages and trade costs. Additionally, idea market access would remain unchanged when the speed of technology diffusion changes, so the asymmetric scale effect cannot be obtained. When innovation and technology diffusion are connected at the level of individual ideas but ideas have identical efficiency across different production locations (Cai et al., 2021; Eaton and Kortum, 2024), colocation and the asymmetric effect can be obtained, but technology diffusion results in large fluctuations in idea market shares and hence trade shares (as seen by setting $\rho = 0$ in equation 19).

Additionally, the markup for each good is the ratio between the second lowest cost (from the second highest quality draw) and the lowest cost (from the highest quality draw). Since idea quality is drawn from a Pareto distribution, the ratio of the second highest quality to the highest quality draw conditional on the second highest quality draw is always Pareto and scale invariant[26]. Consequently, the distribution in which markups are drawn is Pareto and invariant across regions and time. Formally:

Lemma 4. *Given Bertrand competition and the Cobb-Douglas aggregation of goods in each region and sector (equation 16), the* ***markup distribution*** *is Pareto and invariant across idea discovery time t^*, production time t, idea discovery location r, origin o, destination d, and sector k:*

$$G^{(2)/(1)}(m) \equiv \mathbb{P}\left[\frac{C^{(2)}}{C^{(1)}} \leq m | C^{(2)} = c_2\right] = 1 - m^{-\theta}. \tag{25}$$

This distribution is identical to a setting without technology diffusion because I assume that only the greatest application of each idea can be used to produce the corresponding good (see equation 12). Thus, Bertrand competition occurs across ideas, not across applications within the same idea[27]. This invariant markup distribution is crucial for deriving closed-form expressions for aggregate profits and the expected value of individual ideas within each region-sector.

3.3.2 Profits and Innovation Worker Wages

Specifically, the invariant markup distribution implies that all firms in sector k selling in destination d charge a markup drawn from $G^{(2)/(1)}(m)$[28]. Consequently, the total profits earned from selling in destination d, irrespective of the locations of innovation and production, are given by:

$$\Pi_{d,t} = X_{d,t} \int_0^1 1 - \frac{1}{m(\nu)} d\nu = X_{d,t} \int_1^\infty 1 - \frac{1}{m(\nu)} dG^{(2)/(1)}(m) = \frac{X_{d,t}}{1+\theta}, \tag{26}$$

where $X_{d,t}$ is the total spending by production workers and inventors in region d at time t and given by equation (30). Because the markup distribution is identical whether conditional or unconditional on the production location, profits earned by firms from selling their goods can be arbitrarily assigned to production, innovation, or a combination of both[29]. I now make assumptions on how these profits are allocated and subsequently paid to individual inventors.

[26] The scale invariance of the Pareto distribution is *the* fundamental mathematical property that drives the tractability and mathematical beauty of the Eaton-Kortum world.

[27] I make this simplifying assumption because I interpret technology diffusion in my model as primarily occurring *within* firms, to capture the asymmetric scale effect documented in my empirics, where firms initially concentrated in high-skill cities expanded to lower-cost regions disproportionately following the ICT shock. This interpretation is also highly consistent with the stringent intellectual property protection laws in the US. Allowing for both within- and across-firm diffusion, and hence competition between different applications of the same idea, would be more appropriate in an international or developing country setting.

[28] Formally, to derive the unconditional markup distribution across all ideas, we need to integrate the conditional markup distribution $G^{(2)/(1)}(m)$ over all values of the second lowest cost c_2. However, since this conditional distribution is independent of c_2, it is exactly equal to the unconditional markup distribution.

[29] Consequently, my model is consistent with *any* structure of intellectual property protection and/or technology licensing, thereby nesting Hemous et al 2024.

Assumption 2 (Allocation of Profits from Sales)**.**
(i) There is perfect intellectual property protection, such that all profits from sales are allocated to innovation; (ii) At every instant, firms invest their profits in risk-free assets, which are produced using the same technology as final goods in each region[30] *and do not depreciate over time. Firms compensate the inventors they hire with wages equal to the expected return of their innovation efforts.*

Given Assumption 2(i), the expected value of an idea is formalized in the following lemma:

Lemma 5. *Given Bertrand competition, the Cobb-Douglas aggregation across individual goods (equation 16), the time- and region-invariant markup distribution (equation 25), and the allocation of profits from sales (Assumption 2(i)), the* ***expected value of an idea*** *in region r and sector k is:*

$$\check{V}_{r,t^*} = \int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \underbrace{\frac{\phi_{rd,t^*t}}{\lambda_{r,t^*}}}_{\substack{\text{share of profits earned} \\ \text{in region } d \text{ at time } t \\ \text{by an idea discovered} \\ \text{in region } r \text{ at time } t^*}} \cdot \underbrace{\frac{X_{d,t}}{1+\theta}}_{\substack{\text{profits earned} \\ \text{in region } d \\ \text{at time } t \\ \text{by all ideas}}} \cdot \underbrace{\frac{P_{rt^*}}{P_{rt}}}_{\substack{\text{accounting for} \\ \text{changes in} \\ \text{purchasing power} \\ \text{over time}}} dt \tag{27}$$

where θ is the trade elasticity, ζ is the discount rate, and ϕ_{rd,t^*t} is the idea market share.

The expected value of an idea is given by the present discounted value of the trajectory of profits for all $t \geq t^*$ during which the idea is used. The first term captures the probability the idea developed in region r at time t^* results in a good sold in destination market d at time t, the second term represents the total profits earned in destination market d at time t by all ideas, while the third term accounts for changes in purchasing power over time.

Given Assumption 2(ii), wages of inventors are simply the product of the expected number of ideas they discover, as described by the idea production function in equation (36), and the expected value of each idea in equation (5). Formally:

$$w_{r,t^*}^{k,R} = \frac{\lambda_{r,t^*}^{k}}{L_{r,t^*}^{k,R}} \check{V}_{r,t^*}^{k}. \tag{28}$$

3.4 Consumption, Market Clearing, and Production Worker Wages

At each time t, production workers and inventors have identical Cobb-Douglas preferences over local final goods from the ICT and non-ICT sectors, with ι capturing the sectoral expenditure share allocated to goods produced in the ICT sector. Given these preferences, the market clearing condition at each time t is formalized in the following lemma:

[30] This structure mirrors the investment good technology in Kleinman et al. (2023), ensuring that the market clearing condition in each period (see equation 29) is solely determined by contemporaneous variables.

Lemma 6. *Given the allocation of profits from sales in Assumption 2, the combined goods and innovation* ***market clearing*** *condition at time t is given by:*

$$\frac{1+\theta}{\theta} w_{o,t}^k L_{o,t}^k = \sum_d \pi_{od,t}^k \iota^k \left[\sum_s \left(w_{d,t}^s L_{d,t}^s + \sum_r \varphi_{dr,t}^s \frac{1}{\theta} w_{r,t}^s L_{r,t}^s \right) \right] \tag{29}$$

where $\pi_{od,t}^k$ are the trade shares from equation (18), $\varphi_{dr,t}^k$ are the idea adoption shares from equation (22), $w_{o,t}^k$ are the production worker wages, and $L_{o,t}^k$ is the number of production workers.

In particular, total expenditure in region d on sector k goods is given by:

$$X_{d,t}^k = \iota^k \left[\sum_s \left(w_{d,t}^s L_{d,t}^s + \sum_r \varphi_{dr,t}^s \frac{1}{\theta} w_{r,t}^s L_{r,t}^s \right) \right]. \tag{30}$$

Since profits earned from production are a constant multiple of the income earned by production workers in the region-sector, and firms reinvest their profits in the same period to produce assets, neither profits nor wages of innovation workers appear explicitly in the market clearing condition. Consequently, I omitted the superscript G for wages and labor to reduce notational burden.

3.5 Worker Mobility

To allow innovation and production worker wages (equations 28 and 29 respectively) to shape the spatial and sectoral directions of innovation, I introduce worker mobility between production and research as well as across regions and sectors. Worker mobility is also particularly important in a within-country setting[31]. As a middle ground between perfect mobility and immobile workers, I introduce bilateral mobility frictions. To allow shocks to economic fundamentals (such as the ICT shock) to gradually impact the geography of innovation, I model worker migration decisions as dynamic. These dynamic bilateral decisions necessitate additional notation. I denote worker occupations h, n to represent either production (G) or research (R). Since worker mobility in continuous time complicates the distinction between current versus future payoffs, I make the following assumption on the timing of migration:

Assumption 3. *A Poisson arrival process with rate 1 governs when* ***all*** *workers can move.*

In anticipation of when a move arrival occurs at some t', workers decide where to migrate to, which sector to work in (ICT or non-ICT), and whether to engage in production or research based on the present value stream of utility minus the associated mobility costs from their current region, sector, and occupation. With perfect foresight, the optimization problem of a worker in region d, sector k, and occupation h at time t is:

$$v_{d,t}^{k,h} = \max_{o,s,n} \quad \mathbb{E}_t \left(\int_t^{t'} \frac{w_{d,\check{t}}^{k,h}}{P_{d,\check{t}}} d\check{t} \right) + \frac{1}{1+\zeta} \mathbb{E}_t \left(\mathbb{E}_\epsilon \left[v_{o,t'}^{s,n} \right] \right) - \kappa_{do,t}^{ks,hn} + \epsilon_{o,t}^{s,n} \tag{31}$$

[31] In my application, worker mobility across regions captures the underlying trend of workers migrating to high-skill cities documented in Fact 9.

where $\kappa_{do,t}^{ks,hn}$ are the costs of moving from region d, sector k, and occupation h to region o, sector s, and occupation n, ζ is the discount rate, $\epsilon_{o,t}^{s,n}$ is an individual-specific idiosyncratic shock in each potential destination region-sector-occupation, $\mathbb{E}_t(\cdot)$ is the time-t expectation over future state variables, and $\mathbb{E}_\epsilon(\cdot)$ is the expectation over the agent's future realizations of the idiosyncratic shock. At each time t, I assume each individual-specific idiosyncratic shock $\epsilon_{o,t}^{s,n}$ is drawn from a multivariate Gumbel distribution with the following cumulative distribution function:

$$\breve{F}\left(\left\{\epsilon_{o,t}^{s,n}\right\}_{o=1,\ldots,N}^{s=\{\text{ICT,non-ICT}\},n=\{G,R\}}\right)=\exp\left\{-\left[\sum_o\sum_s\left(\sum_n\exp\left(-\epsilon_{o,t}^{s,n}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right]\right\} \tag{32}$$

where υ is the elasticity of worker mobility across regions and sectors, and $\frac{\Upsilon}{\upsilon}$ is the elasticity of worker mobility between production and research.

Lemma 7. *Given individual-level worker mobility decisions defined by equations* (31)-(32)*, the* ***expected value*** *or lifetime utility of a representative worker in labor market* (d,k,h) *is given by:*

$$V_{d,t}^{k,h}\equiv\mathbb{E}_\epsilon\left[v_{d,t}^{k,h}\right]=\mathbb{E}_t\left(\int_t^{t'}\frac{w_{d,\check{t}}^{k,h}}{P_{d,\check{t}}}d\check{t}\right)+\frac{1}{\Upsilon}\log\left[\sum_o\sum_s\left(\sum_n\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n}-\kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right] \tag{33}$$

aggregate ***mobility shares*** *of workers from* (d,k,h) *to* (o,s,n) *is given by:*

$$\begin{aligned}\mu_{do,t}^{ks,hn}\equiv\mathbb{E}_t\left[\check{\mu}_{do,t'}^{ks,hn}\right]&\equiv\mu_{do,t}^{ks,hn}|\mu_{do,t}^{ks}\cdot\mu_{do,t}^{ks}\\&=\underbrace{\frac{\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n}-\kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n'}-\kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}}}_{\textit{switching between production and research}}\cdot\underbrace{\frac{\left[\sum_{n'}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n'}-\kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{o'}\sum_{s'}\left[\sum_{n'}\exp\left(\frac{1}{1+\zeta}V_{o',t'}^{s',n'}-\kappa_{do',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}}_{\textit{migration across regions and sectors}}\end{aligned} \tag{34}$$

and the worker population in (o,s,n) *evolves as follows:*

$$L_{o,t'}^{s,n}=\sum_h\sum_k\sum_d\mu_{do,t}^{ks,hn}L_{d,t}^{k,h}. \tag{35}$$

3.6 Endogenous and Directed Innovation

Up to this point, the Poisson arrival rate of new ideas (or innovation levels) λ_{r,t^*}^k has been treated as exogenous and arbitrary. To introduce endogenous and directed innovation, I let λ_{r,t^*}^k depend on the number of innovation workers in the respective region and sector, ensuring that more innovation workers lead to higher innovation levels. Consequently, as profits and wages from innovation rise in a region, additional workers move into innovation there, endogenously increasing local innovation.

To connect my model with single-region endogenous growth models such as Romer (1990), I assume λ^k_{r,t^*} is a linear function of the region- and sector-specific technology level, T^k_{r,t^*}[32]. Finally, to align my model with the mechanisms through which the ICT shock influenced the geography of U.S. innovation after 1990 – documented in my empirical findings – I allow λ^k_{r,t^*} to also depend on other exogenous and spillover components. Formally, λ^k_{r,t^*} depends on six terms:

$$\lambda^k_{r,t^*} = \underbrace{A_{r,t^*}}_{\substack{\text{fundamental research productivity in region } r \\ \textbf{(function of college ratio)}}} \cdot \underbrace{A^k_{t^*}}_{\substack{\text{sector-specific national research productivity} \\ \textbf{(Direct effect of ICT)}}} \cdot \underbrace{T^k_{r,t^*}}_{\substack{\text{sector-specific technology level in region } r}} \cdot \underbrace{\left(L^{k,G}_{r,t^*}\right)^{\chi}}_{\substack{\text{number of production workers} \\ \textbf{(benefits of colocation with production)}}} \cdot \underbrace{\left(L^{R}_{r,t^*}\right)^{\alpha}}_{\substack{\text{agglomeration economies in innovation including } \textbf{spillovers from ICT to non-ICT}}} \cdot \underbrace{L^{k,R}_{r,t^*}}_{\substack{\text{number of inventors}}} \tag{36}$$

The first term represents the region's fundamental research productivity. It includes a time-invariant component proportional to the region's college ratio – capturing why high-skill cities innovate more – and a time-varying component that accounts for factors unexplained by the model, such as local innovation policies, infrastructure, or other regional trends. The second term denotes the sector-specific research productivity. As described at the beginning of my model, this term incorporates the first component of the ICT shock via a *one-time* increase in ICT research productivity A^{ICT}_{1990}, ensuring that ICT innovation levels rise in all regions relative to a scenario without this shock.

The next two terms capture how the initial concentration of the ICT sector (comprising both innovation and production) interacts with the one-time ICT research shock to intensify the concentration of innovation in high-skill cities. Specifically, the third term represents the region- and sector-specific technology level, showing how a higher initial stock of known ICT production techniques in high-skill cities $T^{ICT}_{r,1990}$, relative to other regions, spurs greater ICT innovation in these cities during the initial period. Through the evolution of technology (equation 15), this effect persists in subsequent periods. The fourth term incorporates the benefits of ***colocation between innovation and production***, modeled as contemporaneous spillovers from the number of production workers in the same region and sector, where $0 < \chi < 1$. This term shows how the initially higher level of ICT employment or production in high-skill cities – relative to other regions – drives greater ICT innovation in these cities during the initial period. As ICT innovation grows, frictions in technology diffusion imply that most of these innovations are adopted locally (see equation 22), raising ICT production worker wages and attracting more workers into ICT production in high-skill cities. This term then captures how increased ICT production fuels further ICT innovation.

The remaining terms account for other mechanisms through which the ICT shock influenced the geography of U.S. innovation. The fifth term represents agglomeration economies in innovation, encompassing ***spillovers from ICT to non-ICT innovation*** (and vice versa). The sixth term is a standard component that denotes how the number of innovation workers in the region and sector directly scales the production of new ideas. In my context, this term captures the impact of the ***asymmetric scale effect*** – arising from reduced

[32] Alternatively, replacing T_{l,t^*} with T^{β}_{l,t^*} with $\beta < 1$ would yield semi-endogenous growth, as in many single-region growth models pioneered by Jones (1995); Kortum (1997).

communication costs (the second component of the ICT shock) – and ***worker migration to high-skill cities*** on the geography of innovation. Specifically, the asymmetric scale effect disproportionately increases the expected value of ideas in high-skill cities (see equation 27), leading to higher inventor wages in these regions (see equation 28). As more workers move into innovation in high-skill cities, this sixth term leads to increased innovation there.

My specification of λ^k_{r,t^*} enables the *one-time* ICT research productivity shock to also capture the nationwide compositional shift in innovation from non-ICT to ICT after 1990 (Fact 2) alongside the rising concentration of innovation in high-skill cities (Fact 1) explained above. Specifically, a one-time increase in A^{ICT}_{1990} raises ICT innovation in the initial period relative to non-ICT across all regions. This, in turn, increases ICT technology levels (see equation 15), ICT innovation workers (driven by higher wages and worker mobility), and ICT production workers (driven by the benefits of colocation and worker mobility) relative to non-ICT in subsequent periods. Together, these dynamics drive a gradual compositional shift of innovation towards ICT via the third, sixth, and fourth terms, respectively.

With the specification of λ^k_{r,t^*}, I now define the equilibrium of my model as follows:

3.7 Definition of Equilibrium

Given an initial distribution of technology levels $\{T^k_{o,0}\}^{N,N}_{o=1,k=1}$ and workers $\{L^{k,h}_{o,0}\}^{N;\{\text{ICT,non-ICT}\};\{G,R\}}_{o=1;k;h}$, trajectories of bilateral trade costs $\{\tau_{od,t}\}^{N,N,\infty}_{o=1,d=1,t=0}$, bilateral migration costs $\{\kappa^{ks,ha}_{od,t}\}^{N;N;\infty;\{\text{ICT,non-ICT}\};\{G,R\}}_{o=1;d=1;t=0;k,s;h,a}$, bilateral diffusion lags $\{\delta_{od,t^*}\}^{N,N,\infty}_{o=1,d=1,t^*=0}$, fundamental productivities in idea production $\{A_{r,t^*}, A^k_{t^*}\}^{N,\{\text{ICT,non-ICT}\},\infty}_{r=1,k,t^*=0}$ and fundamental parameters and elasticities $\{\theta, \sigma, \upsilon, \Upsilon, \iota, \alpha, \zeta\}$, the **dynamic competitive equilibrium** is defined by a trajectory of values, wages, prices and labor allocations $\{V, w, P, L\}$ that satisfy the bilateral migration shares and evolution of worker populations (Lemma 7), innovation levels (equation 36), evolution of technology levels (Lemma 1 equation 15), bilateral trade and idea adoption shares (Lemma 2 equations 18 and 22), price indices (Lemma 3), returns to innovation (Lemma 5) and market clearing condition (Lemma 6).

3.8 Model Interpretation and Extensions

Spatial Direction of Innovation

My model introduces *endogenous and directed innovation in a spatial setting.* To highlight this methodological contribution and connect it to the sectoral direction of technical change developed by Acemoglu (1998, 2002, 2007), I now leverage the expression for inventor wages (equation 28) and my specification of the research production function (equation 36) to define the spatial direction of innovation in my setting, as summarized in the following proposition:

Proposition 1. *The **spatial direction of innovation** is governed by the ratio of sector-specific inventor real wages across regions, given as follows:*

$$\frac{\omega^{k,R}_{r,t^*}}{\omega^{k,R}_{r',t^*}} = \underbrace{\frac{A_{r,t^*}}{A_{r',t^*}}}_{\substack{\text{fundamental research productivity (function of } \textbf{college ratio)}}} \cdot \underbrace{\frac{T^k_{r,t^*}}{T^k_{r',t^*}}}_{\substack{\text{relative technology levels}}} \cdot \underbrace{\left(\frac{L^{k,G}_{r,t^*}}{L^{k,G}_{r',t^*}}\right)^{\chi}}_{\substack{\text{benefits of } \textbf{colocation of innovation and production}}} \cdot \underbrace{\left(\frac{L^{R}_{r,t^*}}{L^{R}_{r',t^*}}\right)^{\alpha}}_{\substack{\text{agglomeration economies in innovation including } \textbf{spillovers from ICT to non-ICT}}} \cdot \underbrace{\frac{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\boldsymbol{\phi_{rd,t^*t}}}{\boldsymbol{\lambda_{r,t^*}}} \frac{Y_{d,t}}{1+\theta} \frac{1}{P_{r,t}} dt}{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\boldsymbol{\phi_{r'd,t^*t}}}{\boldsymbol{\lambda_{r',t^*}}} \frac{Y_{d,t}}{1+\theta} \frac{1}{P_{r',t}} dt}}_{\substack{\text{expected market potential of an idea, with the idea market shares capturing the } \textbf{asymmetric scale effect}}} \tag{37}$$

The higher the ratio in equation (37), the greater the returns to innovation directed toward region r relative to region r'. Within each sector, five main factors shape the incentives to innovate across space: (i) time-varying fundamental research productivity, including a time-invariant component tied to regional college ratios; (ii) relative technology levels; (iii) the benefits of colocation between innovation and production; (iv) agglomeration economies in innovation including spillovers from ICT to non-ICT; and (v) the expected market potential of an idea discovered in that region and sector. The first four factors arise from the research production function governing the Poisson arrival rate of ideas (equation 36), while the fifth factor is the expected value of individual ideas (equation 27). Together, these elements capture the endogenous mechanisms that shape the geography of innovation along the transition path in this model, aligning precisely with my empirical findings. More precisely, any increase in the ratio from equation (37) following the ICT shock – driven by one or more of these mechanisms – causes innovation workers to move from region r' to region r, thereby raising innovation in r relative to r' (due to equation 36).

This ratio serves as an analog to the sectoral direction of technical change, capturing inventor real wages in partial equilibrium to highlight the main forces shaping the spatial direction of innovation. However, it is important to note that inventor real wages are jointly determined in general equilibrium, alongside technology levels, labor allocations, prices, and idea market shares on the right hand side of equation (37). Consequently, inventor real wages – and thus the spatial direction of innovation – are also indirectly influenced by many other factors in the economy, such as production worker wages across all regions. In Appendix C.8, I present the wage ratio that governs the sectoral direction of innovation.

Model Extensions

Additionally, my dynamic spatial model microfounds innovation and technology diffusion *within* the Eaton-Kortum framework. Consequently, it is highly tractable and can accommodate a broad range of extensions. These include dynamic worker sorting, input-output loops, capital accumulation, and the inclusion of amenities with congestion and agglomeration in goods production, as demonstrated in Appendix D.

4 Aggregate Consequences of the Rising Spatial Concentration of Innovation from the ICT Shock

Apart from introducing *endogenous and directed innovation in a spatial setting* to capture the key mechanisms driving the rising concentration of U.S. innovation in high-skill cities, my dynamic spatial model provides a flexible framework to *decompose the welfare impact of any shock to economic fundamentals into its transitory and long-run growth components.* In this section, I develop the methodological tools to obtain this decomposition, which illustrates the aggregate consequences of the ICT shock.

4.1 Balanced Growth Path

I first show that the balanced growth path is highly flexible due to its block-recursive nature. This structure both distinguishes and connects the *causes* and *consequences* of the geography of innovation, while establishing a direct link between quantitative trade and spatial models with innovation and endogenous growth models in macroeconomics. The following proposition formalizes this structure:

Proposition 2. *Along the* ***balanced growth path*** *where all equilibrium variables grow at constant (but possibly different) rates:*
(i) The growth rate of technology in each sector g^k is identical across regions and determined by the solution to the following system of equations:

$$\dot{T}_o^k(t) = \sum_r \gamma_r^k T_r^k(t) \int_{-\infty}^{t} g^k e^{-g^k(t-t^*)} \Omega_{ro}(t-t^*)^{1-\rho} dt^* \tag{38}$$

where γ_r^k is the endogenous region-sector-specific innovation rate. In matrix form, this equation is given by:

$$g\mathbf{T}^k = \mathbf{\Delta}^k(g)\mathbf{T}^k \tag{39}$$

where $\mathbf{T}^k$ is an $N \times 1$ vector with representative element T_o^k and $\mathbf{\Delta}^k(g)$ is an $N \times N$ matrix with representative element:

$$\Delta_{ro}^k(g^k) = \gamma_r^k \int_0^{\infty} g^k e^{-g^k a} \Omega_{ro}(a)^{1-\rho} da.$$

Thus, g^k is the Perron-Frobenius root of equation (39) *with relative technology levels $\mathbf{T}$ corresponding to the Perron-Frobenius eigenvector that is defined up to a scalar multiple;*
(ii) Production worker wages, inventor wages, and the distribution of workers across regions, sectors, and occupations are constant, prices are falling at rate $g_p = \frac{1}{\theta}\sum_k \iota^k g^k$, and the expected value of workers is rising at rate $g_v = \frac{1+\zeta}{\zeta}\frac{1}{\theta}\sum_k \iota^k g^k$.

In terms of aggregate *consequences*, part (i) of the proposition shows how any arbitrary steady-state geography of innovation rates ($\gamma_r^k = \frac{\lambda_r^k}{T_r^k}$) and region-pair idea diffusion speeds (δ_{ro}) can generate a balanced growth path characterized by parallel growth at the sectoral rate g^k, alongside persistent level differences in technology T^k across regions for each sector[33]. Notably, reduced communication costs resulting from the

[33]This result arises because the Poisson arrival rate of ideas depends linearly on technology levels (see equation 36)

ICT shock directly impact aggregate growth, in addition to shaping the geography of innovation. Part (ii) then shows how the sector-specific technology growth rates (g^k) determine the rate at which prices fall and, consequently, the growth rate of workers' expected value on the balanced growth path. This price decline parallels standard macroeconomic growth models. However, unlike those models, the rate at which prices fall here depends on the entire spatial distribution of innovation, rather than solely on the economy's aggregate innovation rate.

In terms of *causes*, part (ii) of the proposition shows that the steady-state geography of innovation rates (γ_r^k) is determined by the steady-state distribution of workers, which in turn is shaped by the exogenous fundamentals of the economy, including bilateral idea diffusion speeds, migration costs, trade costs, and research productivities. The equilibrium expressions for innovation rates, inventor wages and other variables are provided in Appendix C.9. In steady-state, this block of the model mirrors a standard quantitative trade model. Since any arbitrary steady-state geography of innovation rates is consistent with the balanced growth path described above, my model can readily accommodate various mechanisms that drive the geography of innovation in different contexts.

4.2 Transition Path

To better understand how shocks to the economy's fundamentals shape the steady-state distribution of workers and innovation rates, I also characterize the transition path. Specifically, I assume the economy is initially on a balanced growth path[34] and characterize its transition to a new balanced growth path following shocks to fundamentals at time $t = 0$. To align my model with observed data on innovation levels, trade shares, migration shares, and firm establishment networks, and to reduce the data required for simulations, I strengthen Assumption 3 as follows:

Assumption 4. *A Poisson arrival process with rate 1 governs when innovation, production, consumption, and migration occurs across* ***all*** *agents and all regions*[35]*.*

Accordingly, I now interpret λ_{r,t^*}^k as the total number of ideas discovered in region r and sector k during the discrete time period t^*, rather than the Poisson arrival rate of new ideas at the instant t^* in continuous time. Given this assumption, I then extend the **dynamic hat algebra** methodology developed by Caliendo et al. (2019) to simulate the transition path without information on levels of migration and trade costs. Specifically, let $\mathcal{T}$ be the set of times where innovation, production, consumption, and migration occur, $\widetilde{x}_t = x_t e^{-g_x t}$ denote the detrended value with growth rate g_x and $\dot{x}_{t'} = \frac{x_{t'}}{x_t}$ represent changes over time for any variable x, and $u = \exp(V)$. The transition path can then be obtained given the trajectory of time *changes* in migration and trade costs – as opposed to levels – alongside levels of research productivity and diffusion lags. This characterization is formalized in the following proposition:

[34]This assumption substitutes for historical data on innovation levels and diffusion speeds from time $-\infty$ to 0, enabling computation of trade shares at all time periods.

[35]This assumption also ensures that workers' expected values are equalized across regions, sectors, and between production and research when innovation and production occur. An alternative, stronger assumption is that all these activities occur only at the beginning or end of each year. In expectation, both assumptions are equivalent.

Proposition 3. *Given an initial distribution of workers,* ***wages****, technology levels, and migration and trade shares* $\left\{L^k_{o,0}, T^k_{o,0}, \mu^{ks,hn}_{od,0}, \pi^k_{od,0}\right\}^{N;N;\{ICT,non\text{-}ICT\};\{G,R\}}_{o=1;d=1;k,s;h,n}$*, exogenous trajectories of changes in migration and trade costs, research productivity, and bilateral diffusion lags* $\left\{\dot{\kappa}^{ks,hn}_{od,t}, \dot{\tau}_{od,t}, A^k_{r,t}, \delta_{ro,t}\right\}^{N;N;N;\{ICT,non\text{-}ICT\};\{G,R\};\infty}_{o=1;d=1;r=1;k,s;h,n;t=0}$*, the* ***transition path*** *of the economy is characterized by the evolution of the distribution of workers for all* $t^*, t, t' \geq 0$*:*

$$\log\left(\dot{\widetilde{u}}^{k,h}_{d,t}\right) = \log\left(\frac{\dot{w}^{k,h}_{d,t}}{\dot{\widetilde{P}}_{d,t}}\right) + \frac{1}{\Upsilon}\log\left[\sum_n\left[\sum_s\sum_o \mu^{ks,hn}_{do,t'}\left(\dot{\widetilde{u}}^{s,n}_{o,t'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\dot{\kappa}^{ks,hn}_{do,t}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}\right] \tag{40}$$

$$\mu^{ks,hn}_{od,t'} = \frac{\mu^{ks,hn}_{od,t}\left(\dot{\widetilde{u}}^{s,n}_{d,t'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\dot{\kappa}^{ks,hn}_{od,t}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'}\mu^{ks,hn'}_{od,t}\left(\dot{\widetilde{u}}^{s,n'}_{d,t'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\dot{\kappa}^{ks,hn'}_{od,t}\right)^{\frac{\Upsilon}{\upsilon}}} \cdot \frac{\left[\sum_{n'}\mu^{ks,hn'}_{od,t}\left(\dot{\widetilde{u}}^{s,n'}_{d,t'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\dot{\kappa}^{ks,hn'}_{od,t}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{n'}\left[\sum_{d'}\sum_{s'}\mu^{ks',hn'}_{od',t}\left(\dot{\widetilde{u}}^{s',n'}_{d',t'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\dot{\kappa}^{ks',hn'}_{od',t}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}} \tag{41}$$

$$L^{k,h}_{d,t'} = \sum_n\sum_s\sum_o \mu^{ks,hn}_{od,t} L^{s,n}_{o,t} \tag{42}$$

where at each time t *innovation levels are given by equation* (36) *and technology levels by:*

$$\begin{aligned}\lambda^k_{r,t} &= A^k_{t^*}A_{r,t^*}\left(L^{k,G}_{r,t^*}\right)^{\chi}\left(L^R_{r,t^*}\right)^{\alpha}L^{k,R}_{r,t^*}T^k_{r,t^*}\\ T^k_{o,t} &= \sum_{r=1}^{N}\sum_{t^*\in\Im^t_{-\infty}}\Omega_{ro,t^*}(t-t^*)^{1-\rho}\cdot\lambda^k_{r,t^*},\end{aligned} \tag{43}$$

the trade equilibrium is given by:

$$\pi^k_{od,t} = \sum_{r=1}^{N}\sum_{t^*\in\Im^t_{-\infty}} \frac{\Omega_{ro,t^*}(t-t^*)\left(w^{k,G}_{o,t}\tau^k_{od,t}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w^{k,G}_{o',t}\tau^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}} \frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w^{k,G}_{r,t}\tau^k_{ro',t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda^k_{r,t^*}}{\sum_{r'}\sum_{\check{t}\in\Im^t_{-\infty}}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w^{k,G}_{o',t}\tau^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda^k_{r',\check{t}}} \tag{44}$$

$$\varphi^k_{ro,t} = \sum_{d=1}^{N}\sum_{t^*\in\Im^t_{-\infty}} \frac{\Omega_{ro,t^*}(t-t^*)\left(w^{k,G}_{o,t}\tau^k_{od,t}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w^{k,G}_{o',t}\tau^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}} \frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w^{k,G}_{r,t}\tau^k_{ro',t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda^k_{r,t^*}}{\sum_{r'}\sum_{\check{t}\in\Im^t_{-\infty}}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w^{k,G}_{o',t}\tau^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda^k_{r',\check{t}}} \tag{45}$$

$$P^k_{d,t} = \Gamma\left[\sum_{r'=1}^{N}\sum_{t^*\in\Im^t_{-\infty}}\left[\sum_{o'=1}^{N}\Omega_{r'o',t^*}(t-t^*)\left(w^{k,G}_{o',t}\tau^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda^k_{r',t^*}\right]^{-\frac{1}{\theta}} \tag{46}$$

$$\frac{1+\theta}{\theta}w^{k,G}_{o,t}L^{k,G}_{o,t} = \sum_d \pi^k_{od,t}\iota^k\left[\sum_k\left(w^{k,G}_{d,t}L^{k,G}_{d,t} + \sum_r\varphi^k_{dr,t}\frac{1+\theta}{\theta}w^{k,G}_{r,t}L^{k,G}_{r,t}\right)\right]$$

where the market clearing condition comes from equation (29)*, and the wages of inventors, returns to inno-*

vation and the probability that goods sold in destination d at time t' uses ideas discovered in region r at time t are given by:

$$w_{r,t}^{k,R} = \frac{\check{V}_{r,t}^{k} \lambda_{r,t}^{k}}{L_{r,t}^{k,R}} \tag{47}$$

$$\check{V}_{r,t}^{k} = \sum_{t' \in \mathcal{T}_{t}^{\infty}} \left(\frac{1}{1-\rho} \right)^{t'-t} \sum_{d} \frac{X_{d,t'}^{k}}{1+\theta} \cdot \frac{P_{r,t}}{P_{r,t'}} \cdot \frac{\phi_{rd,tt'}^{k}}{\lambda_{l,t}^{k}} \tag{48}$$

$$X_{d,t'}^{k} = \iota^{k} \left[\sum_{k} \left(w_{d,t'}^{k,G} L_{d,t'}^{k,G} + \sum_{l} \varphi_{dr,t'}^{k} \frac{1+\theta}{\theta} w_{r,t'}^{k,G} L_{r,t'}^{k,G} \right) \right] \tag{49}$$

$$\phi_{rd,tt'}^{k} = \frac{\left[\sum_{o'} \Omega_{ro',t}(t'-t) \left(w_{r,t'}^{k,G} \tau_{ro',t'}^{k} \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho} \lambda_{r,t}^{k}}{\sum_{r'} \sum_{\check{t} \in \mathcal{T}_{-\infty}^{t'}} \left[\sum_{o'} \Omega_{r'o',\check{t}}(t'-\check{t}) \left(w_{o',t'}^{k,G} \tau_{o'd,t'}^{k} \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho} \lambda_{r',\check{t}}^{k}}, \tag{50}$$

the growth rate of prices is $g_p = -\frac{1}{\theta} \sum_k \iota^k g^k$ and worker value is $g_v = \frac{1+\zeta}{\zeta} \frac{1}{\theta} \sum_k \iota^k g^k$, with g^k as the Perron-Frobenius root of equation (39)*, as described in Proposition 2.*

The equilibrium conditions on the transition path illustrate the mechanics of innovation in dynamic spatial equilibrium. The evolution of the distribution of workers across regions, sectors, and occupations – given by equations (41) and (42) – along with exogenous research productivities, determine the trajectory of innovation and technology levels given by equations (36) and (43). The trajectory of innovation levels, along with exogenous bilateral diffusion lags, determine the trade equilibrium for each production time, as characterized by equations (44)-(46) and (29). In particular, the market clearing condition pins down contemporaneous production worker wages. The trade equilibrium also yields the probability that goods produced in region o at time t' uses ideas discovered in region r at time t, given by equation (50). Trajectories of this probability determine incentives to innovate, captured by the value of individual ideas in equation (48) and hence inventor wages in equation (47). In turn, the wages of production workers and inventors determine the incentives of workers to migrate – given by equation (40) – and hence determine the evolution of the distribution of workers across regions, sectors, and occupations.

Notice that the trade equilibrium along the transition path (equations 44-46 and 29) cannot be expressed in terms of time differences. Because ideas have varying applicabilities across regions and over time, trade shares depend on the entire history of past innovations from $t = -\infty$, rather than solely on the contemporaneous technology stock. Consequently, two additional sets of variables are required to compute the trade shares and simulate the economy compared to existing quantitative spatial models. First, assuming that the economy is on a balanced growth path for $t \leq 0$, the technology stock at $t = 0$ serves as a sufficient statistic for the trajectory of innovation levels for $t \leq 0$. Second, data on wages at $t = 0$, along with trade shares and technology levels, allow for the imputation of initial trade costs from equation (79). Consequently, simulating the transition path for $t \geq 0$ requires only information on changes in trade costs over time like in standard dynamic spatial models, as these changes can be converted to absolute levels given initial trade costs. Note

also that the trade equilibrium does not incorporate detrended variables because production worker wages remain constant on the balanced growth path and can be determined independently of the sectoral price index, which declines over time. This simplification streamlines the numerical algorithm by allowing prices to be detrended *after* computing the trajectory of trade equilibria[36].

4.3 Welfare Impacts of Shocks to Economic Fundamentals

Given the characterizations of the balanced growth and transition paths, I now decompose the welfare impact of the ICT shock, or any arbitrary anticipated sequence of counterfactual changes in fundamentals into transitory and long-run growth components. Let $\acute{x}$ denote the counterfactual path and $\widehat{x}_{t'} = \frac{\dot{\acute{x}}_{t'}}{\dot{x}_{t'}} = \frac{\acute{x}_{t'}/\acute{x}_t}{x_{t'}/x_t}$ denote counterfactual changes for any variable x. Define the welfare impact in **market** (d,k,h) of an anticipated sequence of counterfactual changes in fundamentals from time $t=0$ as the compensating variation in consumption for market (d,k,h), $\log \delta_d^{k,h}$, given by the following equation:

$$\acute{V}_{d,0}^{k,h} = V_{d,0}^{k,h} + \sum_{t' \in \mathcal{T}_0^\infty} \left(\frac{1}{1+\zeta}\right)^{t'} \log \delta_d^{k,h}.$$

Using these notations and definition of welfare, the impact of an anticipated sequence of counterfactual changes in fundamentals is given by the following corollary:

Proposition 4. *Given the transition and balanced growth paths, the* ***welfare*** *effects in each market (i.e. region-sector-occupation) of an anticipated counterfactual change in fundamentals is given by:*

$$\log\left(\delta_d^{k,h}\right) = \sum_{t' \in \mathcal{T}_{\mathbb{R}+}} \left(\frac{1}{1+\zeta}\right)^{t'} \log\left(\underbrace{\frac{\widehat{w_{d,t}^{k,h}}}{\widehat{\widetilde{P}_{d,t}}}}_{\substack{\text{change in future detrended real wages}}} \underbrace{\frac{1}{\left(\widehat{\mu_{dd,t}^{kk}}\right)^{1/\Upsilon}\left(\widehat{\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}}\right)^{\upsilon/\Upsilon}}}_{\text{change in option value of moving}} \right) + \underbrace{\frac{1}{\theta}\sum_k \iota^k\left(\acute{g}^k - g^k\right)}_{\text{growth effects}}. \quad (51)$$

while local and aggregate welfare is defined as the population-weighed average of the welfare impacts in the relevant markets.

There are two key differences in this welfare expression compared to Caliendo et al. (2019). First, the option value of moving is determined by both the own-migration share across region-sectors and the conditional own-migration share across occupations, because workers' idiosyncratic preferences for each occupation are correlated across region-sectors. Second, the welfare expression accounts for long run growth, since a counterfactual change in fundamentals affects both transitory and growth outcomes in my model.

[36] In a subsequent draft, I will further characterize: (i) the speed of the economy's transition building on tools from Kleinman et al. (2023); (ii) the conditions for the uniqueness of the balanced growth path based on Allen et al. (2023); and (iii) the conditions for the uniqueness of the transition path building on Allen and Donaldson (2022). Notice that my model with endogenous innovation provides a specific microfoundation for spatial path dependence.

5 Conclusion

The rise of high-tech clusters over the past half-century has been a central focus of research on the geography of innovation and a frequent topic in the popular press. Despite extensive attention, the fundamental drivers of this trend have remained elusive, partly due to the wide range of potential explanations.

I tackle this perennial question by leveraging comprehensive data on patents, firms, and inventors from 1976 to 2018 to precisely document when and where innovation became more spatially concentrated. My findings reveal that this rising concentration occurred predominantly in high-skill cities and began only after 1990, suggesting that a significant shock around 1990 triggered the growth of high-tech clusters. Through detailed decompositions and micro-level evidence, I find that the rapid rise of information and communication technologies (the ICT shock) from 1990 explains most of this trend through two distinct channels. First, there was a compositional shift in innovation towards ICT, which is *colocated* with the ICT service sector and concentrated in high-skill cities. Second, firms initially concentrated in high-skill cities produced more non-ICT patents due to *spillovers* from ICT innovation and an *asymmetric scale effect* arising from ICT-enabled reductions in communication costs – these firms disproportionately expanded production to lower-cost regions relative to others, enhancing the profitability of new ideas.

To better understand the central mechanisms – *colocation*, *spillovers*, and the *asymmetric scale effect* – shaping the geography of US innovation following the ICT shock, I develop a model of spatial growth that integrates endogenous innovation with technology diffusion at the level of *individual ideas*, along with dynamic worker mobility with frictions. In my model, the ICT shock is characterized by two exogenous components: a one-time shock to ICT research productivity in the initial period, and a decrease in the national component of bilateral diffusion speeds, reflecting reduced communication costs. My model incorporates *spillovers* from ICT to non-ICT innovation through the research production function and offers analytical characterizations of two key mechanisms: (i) the degree of *colocation* between innovation and production, and; (ii) the *asymmetric scale effect*, whereby a uniform increase in diffusion speeds across all region-pairs disproportionately increases the market access of ideas discovered in high-skill cities. These mechanisms influence the wages of innovation workers and, with worker mobility, shapes the spatial direction of innovation along the transition path following the ICT shock, in line with my empirical findings. The steady state distribution of workers and innovation rates then determine the rate at which prices fall and hence the aggregate growth rate of the economy on the balanced growth path. I use the characterizations of the balanced growth and transition paths to then analytically *decompose the welfare impact of the ICT shock into its transitory and long-run growth components.*

More broadly, my model introduces *endogenous and directed innovation* into existing dynamic spatial models in a highly tractable manner, offering methodological tools to quantify the causes and consequences of the geography of innovation in different contexts. This paper thus establishes the foundation for a broader research agenda on the spatial and network aspects of innovation. Leveraging these methodology tools, my ongoing work includes examining the causes and consequences of the concentration of newer technology vintages in big cities and the rise of cross-region coinventor collaborations since 1976.

References

Acemoglu, D. (1998). Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *The Quarterly Journal of Economics*, 113(4):1055–1089.

Acemoglu, D. (2002). Directed Technical Change. *Review of Economic Studies*, 69(4):781–809.

Acemoglu, D. (2007). Equilibrium Bias of Technology. *Econometrica*, 75(5):1371–1409.

Allen, T. and Arkolakis, C. (2014). Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics*, (2002):1085–1139.

Allen, T., Arkolakis, C., and Li, Xiangliang (2023). On the Equilibrium Properties of Spatial Models. *Working Paper [Forthcoming, AER: Insights]*.

Allen, T. and Donaldson, D. (2022). Persistence and Path Dependence in the Spatial Economy. *Working Paper*, 94(June).

Alvarez, F. and Lucas, R. E. (2007). General equilibrium analysis of the Eaton-Kortum model of international trade. *Journal of Monetary Economics*, 54(6):1726–1768.

Andrews, M. J. and Whalley, A. (2021). 150 Years of the Geography of Innovation. *Regional Science and Urban Economics*, 94(December 2020):103627.

Arcidiacono, P., Bayer, P., Blevins, J. R., and Ellickson, P. B. (2016). Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition. *The Review of Economic Studies*, 83(3):889–931.

Autor, D. H. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5):1553–97.

Ben-Akiva, M. and Francois, B. (1983). Mu-homogenous generalized extreme value model. Technical report, Working paper, Department of Civil Engineering, MIT.

Bernard, A. B., Eaton, J., Jensen, J. B., and Kortum, S. (2003). Plants and Productivity in International Trade. *American Economic Review*, 93(4):1268–1290.

Bikard, M. and Marx, M. (2020). Bridging Academia and Industry: How Geographic Hubs Connect University Science and Corporate Technology. *Management Science*, 66(8):3425–3443.

Buera, F. J. and Oberfield, E. (2020). The Global Diffusion of Ideas. *Econometrica*, 88(1):83–114.

Cai, J., Li, N., and Santacreu, A. M. M. (2021). Knowledge Diffusion, Trade, and Innovation across Countries and Sectors (Working Paper). *SSRN Electronic Journal*.

Cai, S., Parro, F., Caliendo, L., and Xiang, W. (2022). Mechanics of Spatial Growth. *Working Paper*, pages

1–108.

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica*, 87(3):741–835.

Caliendo, L. and Parro, F. (2015). Estimates of the Trade and Welfare Effects of NAFTA. *Review of Economic Studies*, 82(1):1–44.

Choi, K.-H. and Moon, C.-G. (1997). Generalized extreme value model and additively separable generator function. *Journal of Econometrics*, 76(1-2):129–140.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2018). The Geography of Development. *Journal of Political Economy*, 126(3):903–983.

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

Dreisigmeyer, P., Goldschlag, N., Krylova, M., Ouyang, W., and Perlman, E. (2018). Building a Better Bridge: Improving Patent Assignee-Firm Links. *CES Technical Notes Series*, (1).

Eaton, J. and Kortum, S. (2001). Technology, trade, and growth: A unified framework. *European Economic Review*, 45:742–755.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, 70(5):1741–1779.

Eaton, J. and Kortum, S. (2024). Technology and the Global Economy. *Annual Review of Economics [Forthcoming]*.

Eckert, F., Fort, T., Schott, P., and Yang, N. (2020). Imputing Missing Values in the US Census Bureau's County Business Patterns. Technical Report w26632, National Bureau of Economic Research, Cambridge, MA.

Ellison, G. and Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105(5):889–927.

Feldman, M. P. and Kogler, D. F. (2010). *Stylized Facts in the Geography of Innovation*, volume 1. Elsevier B.V.

Fort, T. C., Keller, W., Schott, P. K., Yeaple, S., and Zolas, N. (2020). Colocation of Production and Innovation: Evidence from the United States. *Working Paper*.

Greenstein (2015). *How the Internet Became Commercial: Innovation, Privatization, and the Birth of a New Network*. Princeton University Press.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. *NBER Working Paper*.

Hsieh, C.-T. and Rossi-Hansberg, E. (2021). The Industrial Revolution in Services.

Jiang, X. (2023). Information and Communication Technology and Firm Geographic Expansion. *Working Paper*.

Jones, C. I. (1995). R&D-Based Models of Economic Growth. *Journal of Political Economy*, 103(4):759–784.

Kelly, B., Papanikolaou, D., Seru, A., and Taddy, M. (2021). Measuring Technological Innovation over the Long Run. *American Economic Review: Insights*, 3(3):303–20.

Kerr, W. R. and Fu, S. (2008). The survey of industrial R&D—patent database link project. *The Journal of Technology Transfer*, 33(2):173–186.

Kleinman, B. (2022). Wage Inequality and the Spatial Expansion of Firms. *Working Paper*.

Kleinman, B., Liu, E., and Redding, S. J. (2023). Dynamic Spatial General Equilibrium. *Econometrica*, 91(2):385–424.

Kortum, S. (1997). Research, Patenting, and Technological Change. *Econometrica*, 65(6):1389–1419.

Li, G. C., Lai, R., D'Amour, A., Doolin, D. M., Sun, Y., Torvik, V. I., Yu, A. Z., and Lee, F. (2014). Disambiguation and co-authorship networks of the U.S. patent inventor database (1975-2010). *Research Policy*, 43(6):941–955.

Lind, N. and Ramondo, N. (2023). Global Innovation and Knowledge Diffusion. *American Economic Review: Insights*, 5(4):494–510.

Lind, N. and Ramondo, N. (2024). Global Knowledge and Trade Flows: Theory and Measurement. *Journal of International Economics [forthcoming]*.

Liu, J. (2024). Multinational Production and Innovation in Tandem. *Working Paper*.

McFadden, D. (1978). *Modeling the Choice of Residential Location*, pages 75–96. North Holland, Amsterdam.

Moretti, E. (2013). Real Wage Inequality. *American Economic Journal: Applied Economics*, 5(1):65–103.

Moretti, E. (2021). The Effect of High-Tech Clusters on the Productivity of Top Inventors. *American Economic Review*, 111(10):3328–3375.

Ramondo, N. and Rodríguez-Clare, A. (2013). Trade, Multinational Production, and the Gains from Openness. *Journal of Political Economy*, 121(2):273–322.

Schmoch, U. (2008). Concept of a Technology Classification for Country Comparison. *Final report to the World Intellectual Property Organization*, (June):1–15.

Somale, M. (2021). Comparative Advantage in Innovation and Production. *American Economic Journal: Macroeconomics*.

Xiang, W. (2023). Clean Growth and Environmental Policies in the Global Economy. *Working Paper*.

A Details of Data

Here, I provide detailed information about the data sources I used to measure and document fundamental trends in the geography of innovation in the United States, emphasizing my newly constructed six broad fields from technology classes under the Cooperative Patent Classification (CPC) scheme.

A.1 Patent Data and Cleaning Procedures

Patent data comes from bulk files from the US Patent Trademark Office (USPTO) and US PatentsView (USPV) (https://patentsview.org/). The datasets contain the universe of patents granted in the US (by the USPTO) from 1976-2022. I use only utility patents – which comprise about 95% of all granted patents from 1976 – as they are patents for inventions and provide the most appropriate measure of technological advancements, as opposed to other types of patents such as design, plant, and defensive patents. As explained by the USPTO Patent Technology Monitoring Team (PTMT)[37]:

> *Most analyses of technological activity that incorporate patent data will focus on the activity of utility patents, also known as "patents for inventions". Since design patents are granted for ornamental designs for articles of manufacture and not for inventions, they are usually perceived to have a lesser relationship with technological activity. Similarly, statutory invention registrations and defensive publications do not convey patent protection to disclosed inventions and may have a lesser relationship with technological activity. Plant patents may or may not disclose an invention resulting from technological activity; however, plant patents are numerically small relative to utility patents and are usually handled and analyzed separately.*

Inventors, assignees, and locations are disambiguated by USPV – meaning each inventor, assignee, and city-state pair has a unique ID over time – using the latest machine learning techniques, improving on the algorithms used in Li et al. (2014) and to produce the Connecting Outcome Measures in Entrepreneurship, Technology, and Science (COMETS) database. Cities and states of all inventors living in the US are provided[38]. I use the Google Maps API to geocode all locations and assign them to counties, and standard publicly-available crosswalks to convert counties to commuting zones (CZs) [using Autor and Dorn (2013)], core-based statistical areas (CBSAs), and combined statistical areas (CSAs). The high spatial resolution of inventor locations allows me to conduct my analysis at different geographical scales.

I define the patent year as the application year since that is the closest to when the invention was produced, as there are often lags of several years between when a patent was applied and when it was granted. I assign patent shares, citations made, and citations received equally across all coinventors on a patent. I keep only patents where at least one inventor lives in the US.

[37] The USPTO PTMT explanation on the conventions of treating different types of patents can be found at https://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm. Should this or any of the subsequent web links become inactive, PDF copies of the contents of the archived websites that includes the date of access will be provided upon request.

[38] Prior studies have typically only used the state of the first inventor on patents.

A.2 Classification of Patents into Technology Fields and Subfields

Patents are assigned to 3-digit technology classes, 4-digit subclasses, and further divided into groups within subclasses under the International Patent Classification (IPC) and Cooperative Patent Classification (CPC) systems. The IPC provides a hierarchical system of language independent symbols for the classification of patents into eight sections with approximately 70,000 subdivisions. The IPC was developed under the 1971 Strasbourg Agreement and provides a common classification for patents filed in different patent offices around the world. The CPC is a unified system developed jointly by the United States Patent Trademark Office (USPTO) and European Patent Office (EPO) on 2010 to provide a common, internationally compatible classification system that provides more groups and subgroups relative to the IPC. For patents granted in the US, the CPC supercedes the US Patent Classification (USPC) developed by the USPTO and the NBER patent classification developed by Hall et al. (2001). The USPTO no longer provides the USPC and NBER patent classifications for patents granted after 2015[39].

Most patents have more than one field classification to facilitate easier searches to prior art. Nonetheless, under the CPC, each patent has a unique primary field classification, denoted as the "first" position[40]. Other classifications are denoted as having a "later" position. USPTO Guideline 905.03(a)III.A.(C)[41] states:

> *There is one and only one "first" position attribute per patent family. The first attribute is associated with the invention symbol that most completely covers the technical subject matter of the disclosed invention. The first position symbol is identified as the first mandatory symbol listed on the classification form.*

The USPTO provides a bulk file of all granted patents from 1790 with their current CPC classes and position of each CPC class[42]. I use the 08/02/2022 version, downloaded on 09/09/2022[43]. I use the primary CPC class of each patent. In rare cases where there are multiple CPC classes listed as the "first" position, I use the class that is listed first. In rare cases where the CPC class listed in the "first" position is only meant as an additional classification as noted under the details of the CPC class in their classification documentation[44], I use the class in the "later" position that is listed first.

Under the IPC and CPC, the technology subclasses are grouped into 8 broad sections. These broader sections facilitate the allocation of patents to different examiner units at patent offices. **However, some sections contain patents from highly disparate economic fields. For example, Section A includes patents for medical technologies and amusement parks.** To provide more consistent categories with similar sizes, the World Intellectural Property Organization (WIPO) developed a mapping from the 4-digit technol-

[39] Most papers in empirical innovation use the superceded USPC classification and focus on patents granted before 2015.

[40] Several prior studies have incorrectly claimed that there is no primary field classification under the CPC.

[41] The USPTO guidelines on the CPC scheme can be found at https://www.uspto.gov/web/offices/pac/mpep/s905.html

[42] The bulk files containing the CPC classification of every patent can be found at https://bulkdata.uspto.gov/data/patent/classification/cpc/

[43] Although USPV provides CPC classes and technology fields for most patents granted since 1976, data checks I conducted in Summer 2022 indicate that the 2022 vintages of the USPV datasets contain errors in the CPC classes and WIPO technology fields for approximately 3–5% of patents.

[44] The CPC classification documentation can be found at https://www.uspto.gov/web/patents/classification/cpc/html/cpc.html.

ogy subclasses and groups in the IPC/CPC to 35 fields (Schmoch, 2008). This WIPO report also provides 6 broad categories, but the broad category "Instruments" is difficult to interpret and the "Electrical Engineering" category includes fields in information technology as well as older electrical machinery. Thus, I provide an alternative mapping of these 35 WIPO fields into 6 broad categories: (1) Physics, Electrical Engineering & Electronics; (2) Information Technology; (3) Chemistry; (4) Biology & Medicine; (5) Mechanical Engineering; (6) Civil Engineering & Consumer Goods. These categories are similar in scope to the older field classifications in the NBER and COMETS datasets and may be seen as an updated version of them.

The IPC/CPC class to WIPO field concordance is primarily at the level of 4-digit technology subclasses, but there are 6 subclasses with different groups that map to different WIPO fields – A61K, B01D, C13B, E01F, G01N, H04N – due to significantly different subject matter across groups within these subclasses. For example, within the subclass A61K, there are patents for cosmetics and patents for medical technologies. Thus, I provide a further decomposition for these 6 subclasses that corresponds to the WIPO field they are assigned to. My added subclasses are: A61K-14; A61K-16; B01D-23; B01D-24; C13B-18; C13B-29; E01F-24; E01F-35; G01N-10; G01N-11; H04N-2; H04N-3; H04N-4. The numbers after the hyphen refer to the WIPO fields that groups within each subclass are assigned to. Thus, my data has 586 technology classes, comprising the 579 CPC/IPC subclasses along with this decomposition of 6 subclasses. In this paper, I refer to the 35 World Intellectural Property Organization (WIPO) fields as "subfields", my 6 broad categories as "fields", and the 579 CPC/IPC subclasses and 7 additional group categories as "classes".

A.3 Data Sources for Regional Outcomes

I use data on county-level educational attainment for each decade from the Economic Research Service (ERS) of the US Department of Agriculture (USDA) to construct the college ratio and percent of workers that are college educated for each CZ in 1970, 1980, 1990, 2000, 2010 and 2017. The 1970, 1980, 1990 and 2000 data are constructed from the Decennial Censuses of Population and equivalent to the 5% samples from the IPUMS used by Moretti (2013); Diamond (2016) to construct college percent and ratio respectively for metropolitan statistical areas (MSAs) in 1980 and 2000. The 2010 and 2017 data are constructed from 5-year averages from the annual American Community Survey (ACS).

I obtain annual county-level employment by NAICS industry for 1975-2018 from the harmonized County Business Patterns (CBP) Database produced by Eckert et al. (2020). They develop a linear programming method to impute suppressed county-industry-year cells in raw CBP files released by the US Census Bureau and show that total non-agricultural employment from their dataset is highly correlated with the panel from the Longitudinal Business Database (LBD). I use their county-industry employment panel to construct my primary measures for the annual industrial composition of each CZ and the United States, such as the employment share in the ICT service sector. Building on Fort et al. (2020), I define the ICT service sector to include the following industries in the Information Sector (NAICS 51): Software Publishers (5112); Telecommunications (517); Data Processing, Hosting, and Related Services (518).

I also obtain annual county-level data on income and resident population for 1969-2018 from the BEA to

construct basic annual measures of each CZ such as income per capita and population density.

B Additional Empirical Results

B.1 Additional Details on Fact 1 (Rising Concentration in High-Skill CZs from 1990)

Figure 16 presents robustness checks on the trend in the aggregate spatial concentration of innovation in Figure 1 in the main paper. The spatial concentration of innovation only started rising in 1990, whether I drop top patenting CZs (left graph) or use alternative measures of spatial concentration such as the Herfindahl index, coefficient of variation, simplified Ellison-Glaeser measure of patent shares (middle graph), or the annual share of patents produced by the top 10 CZs (right graph).

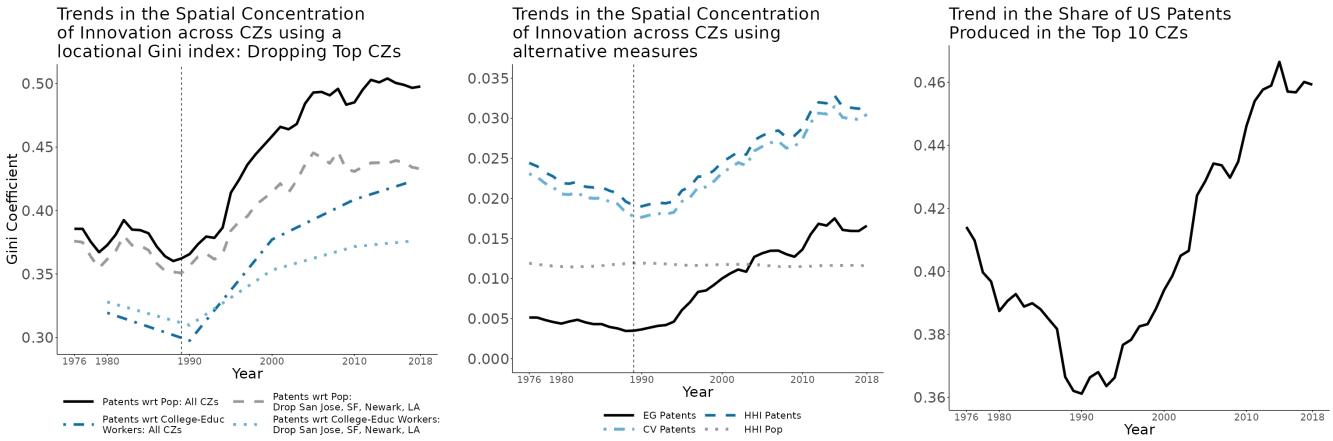


Figure 16: Robustness of trends in the spatial concentration of innovation across CZs. The top left graph plots trends in the locational Gini index of patents with respect to population and college-educated workers after removing top CZs in innovation: San Jose, San Francisco, Newark, and Los Angeles. The middle graph plots trends in alternative measures of spatial concentration, such as the Herfindahl index (blue dashed line), coefficient of variation (light blue dotdashed line), and simplified Ellison-Glaeser measure (black solid line) of patent shares. The right graph plots trends in the annual share of patents produced by the top 10 CZs.

Table 1 lists the top 15 CZs by changes in patent-to-population share between 5-year averages around 1990 and 2015. Apart from well-known superstar cities such as San Jose, San Franciso, San Diego, Seattle and Boston, CZs like Portland, Boise, Wayne, Provo and Fort Collins also became some of the most innovative regions in the US from 1990-2015. Despite narratives of the apparent collapse of innovation in Rochester in several papers and the popular press, the region actually experienced the second greatest increase in patenting intensity from 1990-2015.

Figure 17 plots trends in the annual elasticity of CZ patents per capita with respect to the 1990 population (left graph) and population density (right graph). The left graph shows that the annual elasticity of patents per capita with respect to the 1990 population increased moderately from 1990-2018, these increases did not overcome the decrease in annual elasticity from 1976-1990. The right graph shows that annual elasticity of patents per capita with respect to the 1990 population density did not increase after 1990.

	CZ	Change in patent-to-population share 1990-2015	Patent-to-population share	Rank of patent-to-population share	Patent Share	Patents	Population
1	San Jose	9.494	12.421	1	0.104	14581.5	2684061
2	Rochester	3.608	4.618	2	0.004	531.4	263093
3	San Francisco	3.029	4.440	3	0.073	10153.8	5229363
4	San Diego	2.094	3.430	5	0.035	4882.0	3254818
5	Seattle	2.063	2.955	8	0.043	5974.0	4622131
6	Portland	1.618	2.523	11	0.018	2521.7	2285666
7	Boise	1.411	1.962	18	0.004	604.7	704711
8	Raleigh	1.398	2.282	13	0.015	2064.7	2068936
9	Wayne	1.396	2.303	12	0.001	132.5	131541
10	Austin	1.258	2.959	7	0.019	2660.2	2055406
11	Provo	1.232	1.763	19	0.003	467.1	605648
12	Bloomington	1.186	1.587	25	0.001	157.2	226557
13	Burlington	1.167	2.224	14	0.002	333.9	343199
14	Boston	0.885	2.730	9	0.046	6488.4	5434761
15	Fort Collins	0.871	1.756	20	0.004	495.9	645656
	Oneonta	-0.944	4.414	4	0.002	300.3	155554
	Albany	0.466	2.959	6	0.010	1461.1	1128888
	Elmira	0.341	2.627	10	0.003	401.9	349814
	Poughkeepsie	0.554	2.185	15	0.006	897.9	939410
	Minneapolis	0.192	2.021	16	0.021	2989.4	3381353
	Detroit	0.588	1.967	17	0.032	4510.1	5243719

Table 1: The top 15 CZs by change in ratio of patent-to-population share from 1990-2015 with a population of at least 100,000 in 2015. To minimize the effect of idiosyncratic fluctuations, I take five-year averages around 1990 and 2015. I append this list with the 6 CZs that are in the top 20 by ratio of patent-to-population share in 2015 but did not experience the greatest increase from 1990 to 2015.

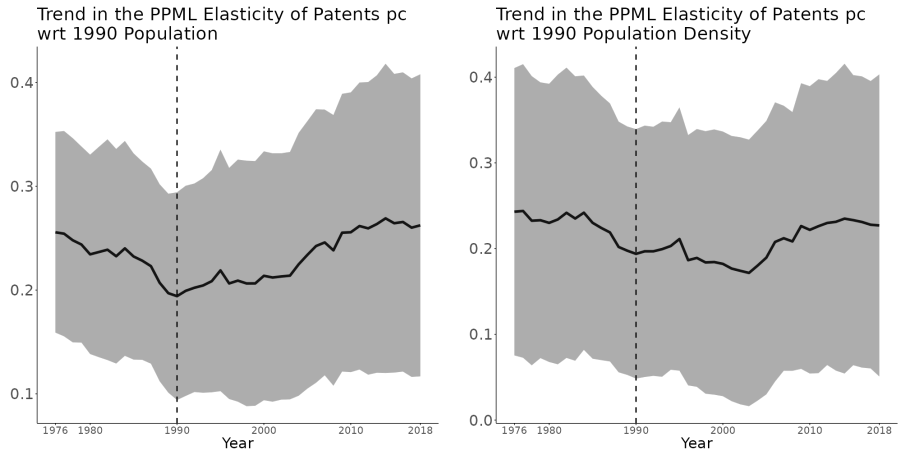


Figure 17: Trends in the annual elasticity of CZ patents per capita with respect to the 1990 population (left) and 1990 population density (right).

B.2 Additional Details on Fact 2 (the ICT Shock): History of the NSFNET

The National Science Foundation Network (NSFNET), developed between 1986 and 1995, was created to facilitate collaboration among researchers at universities and military bases across the United States. It eventually became a crucial bridge between ARPANET – the first public packet-switched computer network operated by the Defense Advanced Research Projects Agency from 1969 to 1989 – and the commercial networks that began providing internet access to the public, particularly during the internet boom starting in 1995.

The history of the NSFNET is defined by five key events:

- Initial establishment: In 1985, the NSF funded the creation of five supercomputing centers. In 1986, it established a long-haul backbone network with a data speed of 56 Kbps, connecting these new centers to the existing supercomputing facility at the National Center for Atmospheric Research.

- First round of upgrading: On June 15, 1987, the NSF issued a solicitation to upgrade and expand the backbone network, addressing the overwhelming demand that had saturated the existing infrastructure. On November 24, 1987, a contract was awarded to a team comprising IBM, MCI, Merit, and the University of Michigan. By July 1988, the upgraded T1 backbone was completed, increasing the number of backbone sites from 6 to 13 and raising network speeds to 1.5 Mbps. The upgraded backbone also enabled connections to regional and campus networks. Each partner played a specific role: Merit developed user support and information services, IBM provided hardware, software, and network management tools, and MCI supplied transmission circuits with reduced tariffs.

- Second round of upgrading: In 1991, the NSF completed the upgrade from the T1 network to a T3 network, increasing the broadband speed to 45 Mbps and adding three new backbone sites.

- **Privatization**: Discussion about privatization began quietly in 1989 and become public by 1990. **In March 1991, the Acceptable Use Policy – which previously required the network to be used solely for research and education – was revised to allow private users**. In 1992, the NSF announced plans to decommission the NSFNET by 1995. By May 1993, it issued a solicitation to encourage more companies to contribute to the development of the Internet's privatized structure[45]

- Decommission: 1995

The most significant event in the history of the NSFNET for U.S. patenting – predominantly conducted by private firms – was the modification of the Acceptable Use Policy in March 1991. This change granted firms access to the NSFNET, significantly reducing their communication costs. Table 2 lists the NSFNET nodes, including the year each was established and the commuting (CZ) in which it is located.

[45]Taken from Kesan, Jay P., and Rajiv C. Shah. 2001. "Fool Us Once, Shame on You—Fool Us Twice, Shame on Us: What We Can Learn from the Privatizations of the Internet Backbone Network and the Domain Name System." Washington University Law Quarterly 79: 89–220.

S/N	NSFNET Node Location	CZ	CZ Name	Year	Selected Universities
1	John von Neumann Center in Princeton, NJ	19600	Newark	1986	Princeton University
2	Cornell Theory Center in Ithaca, NY	18100	Elmira	1986	Cornell University
3	Pittsburgh Supercomputing Center in Pittsburgh, PA	16300	Pittsburgh	1986	Carnegie Mellon University, University of Pittsburgh
4	San Diego Supercomputer Center in San Diego, CA	38000	San Diego	1986	UC San Diego
5	National Center for Supercomputing Applications in Urbana, IL	23500	Decatur	1986	University of Illinois, Urbana-Champaign
6	National Center for Atmospheric Research in Boulder,CO	28900	Denver	1986	University of Colorado Boulder, Colorado School of Mines
7	Palo Alto, CA	37500	San Jose	1988	Stanford University
8	Houston, TX	32000	Houston	1988	Rice University
9	Ann Arbor, MI	11600	Detroit	1988	University of Michigan at Ann Arbor
10	College Park, Maryland	11304	Washington DC	1988	Georgetown University, University of Maryland
11	Salt Lake City, UT	36100	Salt Lake City	1988	
12	Seattle, WA	39400	Seattle	1988	
13	Lincoln, NE	28101	Lincoln	1988	
14	Cambridge, MA	20500	Boston	1991	Harvard University, MIT, Tufts University
15	Argonne National Laboratory in Lemont, IL	24300	Chicago	1991	University of Chicago, Northwestern University
16	Atlanta, Georgia	9100	Altanta	1991	Emory University

Table 2: List of NSFNET backbone nodes

B.3 Additional Details on Fact 3 (Field Decomposition)

Here I present results from my decomposition of the Gini coefficient, as well as robustness on most results when dropping top patenting CZs.

B.3.1 Decomposition of the Rising Gini Coefficient of Patents per capita from 1990 into Within-Field and Cross-Field Components

I decompose changes over time in the overall Gini coefficient G of patents per capita from 1990-2018 into within-field, cross-field, and field colocation components:

$$G_{t^*} - G_{1990} = \sum_{t=1991}^{t=t^*} \Delta G_t = \sum_{t=1991}^{t=t^*} \Bigg[\underbrace{\sum_f \overline{G}_{f,t} \Delta s_{f,t}}_{\substack{\text{changes in field} \\ \text{composition}}} + \underbrace{\sum_f \overline{s}_{f,t} \Delta G_{f,t}}_{\substack{\text{within-field} \\ \text{changes}}} + \underbrace{\Delta \Big(G_t - \sum_f s_{f,t} G_{f,t} \Big)}_{\substack{\text{changes in the} \\ \text{colocation of fields}}} \Bigg]$$

where $G_{f,t}$ is the Gini coefficient of patents in field f with respect to population across CZs in year t, $s_{f,t}$ is the share of US patents in field f in year t, and $\overline{x}_t = \frac{x_t + x_{t-1}}{2}$, $\Delta x_t = x_t - x_{t-1}$ for any variable x. The first

term captures the role of changes in the field composition of US patents. This term is positive if patents are increasingly produced in fields that are more spatially concentrated. The second term captures the role of changes in the spatial concentration of patents within fields. The third term captures the role of changes in the colocation of fields, measured by differences in the overall locational Gini coefficient of all CZ patents with respect to CZ population from a weighted mean of the field-specific locational Gini coefficients.

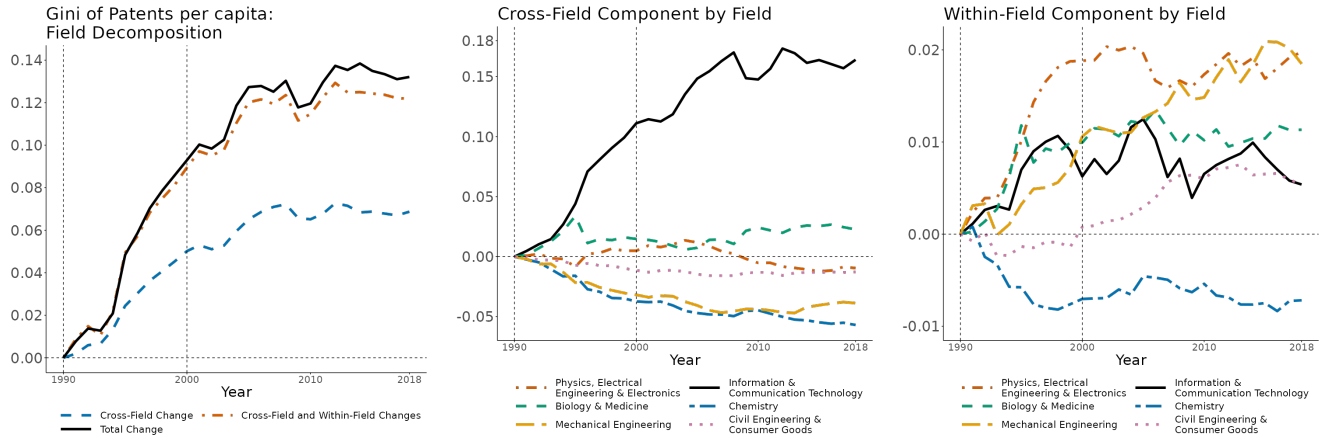


Figure 18: Trends in the decomposition of changes in the locational Gini coefficient of patents with respect to population relative to 1990. The left graph plots trends in the overall decomposition into within-field, across-field, and field colocation components while the middle and right graphs further decompose the cross-field and within-field components respectively into the contribution of each individual field.

The left graph in Figure 18 plots trends in this decomposition: from 1990-2018, 52% of the rise in the spatial concentration of innovation is driven by the changing field composition of patents, 41% by the rising spatial concentration within fields, and 6% by the rising colocation of patents in different fields. The right graph further decomposes the cross-field component into the contribution of each individual field and shows that virtually all of the cross-field component is explained by the rising share of ICT patents.

Figure 19 breaks down ICT into its component subfields, and shows that most of the rising share of ICT patents is driven by Digital Communications and Computer Technology. The top graphs decompose trends in the changes in the aggregate locational Gini coefficient of CZ patents with respect to population into cross-subfield and within-subfield components (left) and the six leading subfields within the cross-subfield component (right). The bottom graphs plot trends in the annual share of patents and the locational Gini coefficient for these six subfields. These graphs show that the rising annual share of patents in Digital Communication and Computer Technology accounts for most of the effects of the rising share of ICT patents on the spatial concentration of innovation.

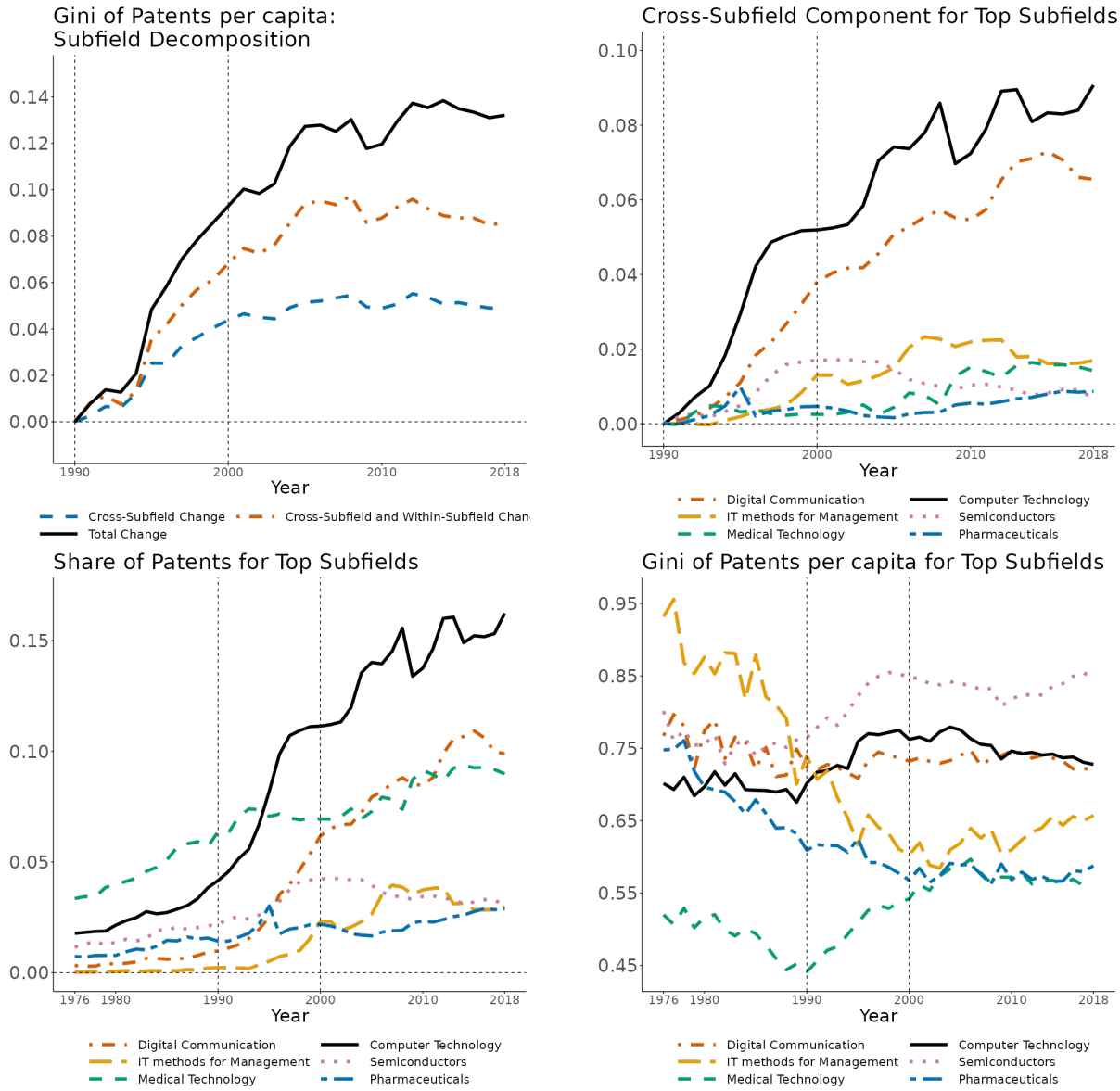


Figure 19: Trends in the subfield decomposition of changes in the aggregate locational Gini coefficient of patents with respect to population. The top left graph plots trends in the decomposition of the aggregate locational Gini coefficient into cross-subfield and within-subfield components. The top right graph plots trends in the top six subfields by contribution to the overall cross-subfield component from 1990-2018. The bottom left graph plots the annual share of patents in these six subfields. The bottom right graph plots the locational Gini coefficient of CZ patents with respect to population for these six subfields.

B.3.2 Robustness of Trends to Dropping Top Patenting CZs

Figure 20 shows that this rise is not just driven by the ICT hub in Silicon Valley. The top graphs plot trends in the annual share of patents (left) and the locational Gini coefficient of CZ patents with respect to population (right) by field – analogous to Figure 7 in the main paper – dropping San Jose and San Francisco. The bottom graphs plot trends in the decomposition of the Gini coefficient of CZ patents with respect to population (left) and a further decomposition of the cross-field component by field (right) – analogous to

Figure 18 – dropping San Jose and San Francisco. After dropping all patents produced in San Jose and SF, 66% of the overall increase (as opposed to 52% with San Jose and San Francisco) in the spatial concentration of innovation is driven by the rise in ICT patents. Thus, these graphs show that the rising annual share of ICT patents can explain a large percentage of the rising spatial concentration of innovation, even after dropping the primary ICT hub in Silicon Valley.

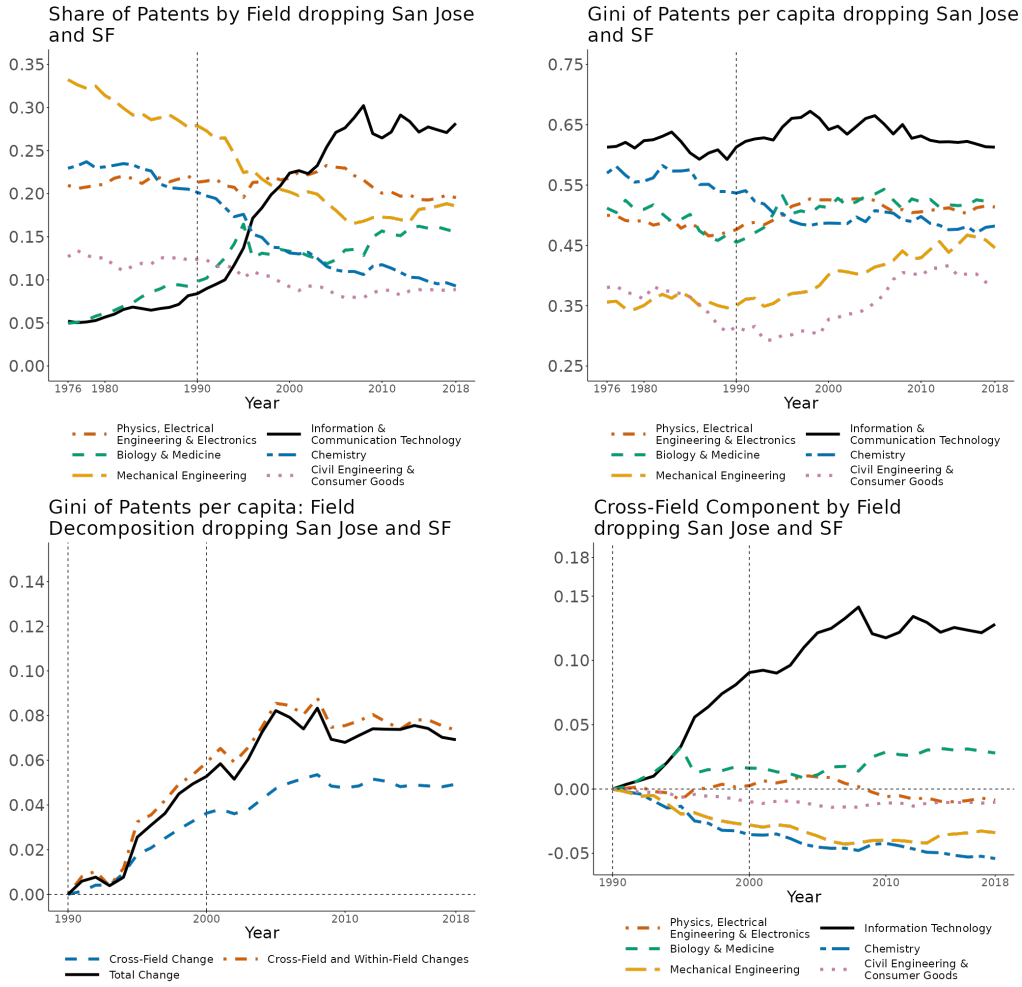


Figure 20: Trends in the share of annual patents by field (top left), the Gini coefficient of CZ patents with respect to CZ population by field (top right), decomposition of the overall Gini coefficient into within-field and cross-field components (bottom left), and a further decomposition of the cross-field component into each individual field (bottom right) after dropping San Jose and San Francisco.

C Proofs and Extensions of Propositions and Lemmas

Lemmas 1 to 6 provide the equilibrium conditions that govern endogenous innovation and technology diffusion; Lemma 7 details the equilibrium conditions of dynamic worker mobility with frictions; Proposition 1 characterizes the *spatial direction of innovation* and; Propositions 2 to 4 *decomposes the welfare impacts of any shock to economic fundamentals into its transitory and long-run growth components.*

Specifically, in the proofs of Lemmas 1 and 2, I leverage techniques from Eaton and Kortum (2024) to provide a more general derivation of the multivariate productivity distribution, minimum cost distribution, and trade shares relative to Lind and Ramondo (2024). This generalization allows me to introduce imperfect competition, characterized by the price index (Lemma 3) and markup distribution (Lemma 4). Along with worker mobility (Lemma 7), I thus *introduce directed innovation in a dynamic spatial model*, characterized by the ratio of inventor wages across regions (Proposition 1) which incorporate the expected value of individual ideas (Lemma 5).

C.1 Lemma 1 (Productivity Distribution)

Proof. I first derive the conditional distribution of applicabilities of an idea, then the goods productivity distribution across all ideas.

(i) Distribution of Applicabilities of an Idea

Microfoundations After an idea is discovered in region r at time t^*, independent stochastic processes of rate $\frac{\tilde{\Omega}}{t-t^*} \cdot \Omega'_{ro,t^*}(t-t^*)dt$ govern its diffusion to each of the other regions $o \neq r$. $\Omega'_{ro,t^*}(t-t^*)$ captures how the density or fertility of an idea in creating applications evolves with the age of the idea, $t-t^*$.

At each instance the idea arrives in region o, its applicability a is stochastic and drawn from a Pareto distribution:

$$H^A(a) = 1 - \left(\frac{a}{\underline{a}}\right)^{-\sigma}, \qquad a \geq \underline{a}, \quad \sigma > \theta, \quad \underline{a} = \tilde{a} \cdot \Gamma\left(1 - \frac{\theta}{\sigma}\right)^{-\frac{1}{\theta}}.$$

Thus, the probability that each application is greater than a is:

$$p_a = 1 - H^A(a) = \Gamma\left(1 - \frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} a^{-\sigma}.$$

The number of applications of this idea in region o that is greater than a by time t follows a binomial distribution with expectation:

$$\mathbb{E} = \Omega_{ro,t^*}(t-t^*) \cdot \Gamma\left(1 - \frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \cdot a^{-\sigma} \cdot \tilde{\Omega}\tilde{a}^{\sigma}.$$

I assume that as the number of applications get arbitrarily large ($\tilde{\Omega} \to \infty$) and the applicability of a typical idea becomes arbitrarily small ($\tilde{a}^{\sigma} \to 0$), its product converges to 1 ($\tilde{\Omega} \cdot \tilde{a}^{\sigma} \to 1$). In this limit, the number

of applications greater than a by time t converges to a Poisson distribution with the same expectation. Thus, the number of applications with applicability $A > a$ that have arrived in region o by time t of an idea discovered in region r at time t^* is distributed Poisson with parameter:

$$\lambda^A(a) = M_{ro,tt^*} a^{-\sigma}, \qquad M_{ro,tt^*} \equiv \Omega_{ro,t^*}(t-t^*) \cdot \Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}}, \quad a > 0,$$

which corresponds to equation (10) in the main text. The expected number of applications of this idea with applicability above a to ever arrive in region o is $\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} a^{-\sigma}$. The expected share of these applications that arrives by time t is $\Omega_{ro,t^*}(t-t^*)$.

Distribution of Applicabilities Using this Poisson distribution, the distribution of the best applicability of this idea in region o by time t is given by the probability that there is no application more efficient than a:

$$F^A_{ro,t^*t}(a) = \mathbb{P}\left[A_o \leq a\right] = e^{-M_{ro,tt^*} a^{-\sigma}} = \exp\left[-\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \Omega_{ro,t^*}(t-t^*) \cdot a^{-\sigma}\right], \tag{52}$$

which is a Frechet distribution. Across all regions, the joint distribution of the best application in time t of an idea discovered in region r at time t^* is:

$$\begin{aligned} F^A_{r,t^*t}(a_1,\ldots,a_N) &= \mathbb{P}\left[A_1 \leq a_1,\ldots,A_N \leq a_N\right] = e^{-\sum_o M_{ro,tt^*} a_o^{-\sigma}} \\ &= \exp\left[-\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \sum_o \Omega_{ro,t^*}(t-t^*) \cdot a_o^{-\sigma}\right], \end{aligned} \tag{53}$$

which is the product of independent Frechet distributions in each region o.

(ii) Goods Productivity Distribution

Microfoundations In region r at time t^*, ideas are discovered from a stochastic process at rate $\lambda_{r,t^*}dt$. At discovery, each idea i has stochastic quality drawn from a Pareto distribution:

$$H^Q(q) = 1 - \left(\frac{q}{\underline{q}}\right)^{-\theta}, \qquad q \geq \underline{q}, \quad \theta > 0,$$

and pertains to the production of a specific good drawn from the unit interval. I assume that as the number of ideas gets arbitrarily large ($\int_0^{t^*} \lambda_{r,t}dt \to \infty$), such that their quality converges to 0 ($\underline{q}^\theta \to 0$), their product $\underline{q}^\theta \int_0^{t^*} \lambda_{r,t}dt$ remains constant. Following a similar argument for the arrival of idea applications above, the number of ideas with quality greater than q that have been discovered in region r by time t^* is thus distributed Poisson with parameter:

$$\lambda^Q_{r,t}(q) = q^{-\theta}\underline{q}^\theta \int_0^{t^*} \lambda_{r,t}dt,$$

which corresponds to equation (9) in the main text.

After an idea is discovered in region r at time t^*, its diffusion to each of the other regions $o \neq r$ is described in part (i) above. The efficiency of an idea i to produce the corresponding good ν in region o at time t is the product of its quality and best applicability in region o at time t:

$$z_i = q_i a^*_{i,o,t}, \tag{54}$$

where $a^*_{i,o,t}$ is drawn from equation (52). Each region o produces each good ν with the most efficient idea:

$$Z_{o,t} = \max_i \left\{ q_i a^*_{i,o,t} \right\}. \tag{55}$$

Marginal Productivity Distribution In Each Region Using the definition of idea efficiency in equation 54, I derive the Poisson process of idea application arrivals in region o incorporating both idea quality on its discovery in region r and idea applicability on its arrival in region o. The expected number of ideas with efficiency $Z > z$ in region o at time t is the total number of ideas produced across all regions with quality above $q = \frac{z}{A_{o,t}}$, where $A_{o,t}$ is the best application of the idea in region o by time t and held fixed when computing q:

$$\begin{aligned}
\lambda^Z_{o,t}(z) &= z^{-\theta} \sum_{r=1}^{N} \int_0^t \int_0^\infty A^\theta dF^A_{ro,tt^*}(a)\ \lambda_{r,t^*} dt^* \\
&= z^{-\theta} \sum_{r=1}^{N} \int_0^t \Gamma\left(1 - \frac{\theta}{\sigma}\right) \cdot \left[\Gamma\left(1 - \frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \Omega_{ro,t^*}(t - t^*) \right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*} dt^* \\
&= z^{-\theta} \sum_{r=1}^{N} \int_0^t \left[\Omega_{ro,t^*}(t - t^*)\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*} dt^*
\end{aligned} \tag{56}$$

where the second equality comes from the mean of the Frechet distribution $F^A_{ro,tt^*}(a)$ in equation (52). Thus the total number of ideas with efficiency above z in region o at time t is distributed Poisson with parameter $\lambda^Z_{o,t}(z)$:

$$P\left[I = i\right] = e^{-\lambda^Z_{o,t}(z)} \frac{\left[\lambda^Z_{o,t}(z)\right]^i}{i!}.$$

The probability that the i'th most efficient idea has efficiency below z is the probability that at most $i - 1$ ideas have efficiency greater than z:

$$F^{Z(i)}_{o,t}(z) = e^{-\lambda^Z_{o,t}(z)} \sum_{l=0}^{i-1} \frac{\left[\lambda^Z_{o,t}(z)\right]^l}{l!}.$$

The marginal productivity distribution in region o at time t is given by the distribution of the most efficient idea:

$$F^{Z(1)}_{o,t}(z) = \mathbb{P}\left[Z^{(1)}_{o,t} \leq z\right] = \exp\left[-\lambda^Z_{o,t}(z)\right] = \exp\left[-T_{o,t} z^{-\theta}\right]$$

where:

$$T_{o,t} = \sum_{r=1}^{N} \int_0^t \left[\Omega_{ro,t^*}(t-t^*)\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*} dt^*$$

which correspond to equations (14) and (15) in the main text.

Joint Productivity Distribution I now derive the Poisson process for idea application arrivals across all regions incorporating quality and applicability, and consequently the joint productivity distribution. The expected number of ideas with efficiency above z_o in at least one region at time t is equivalent to the expected number of ideas with quality $q > \min_o \frac{z_o}{A_{o,t}}$, where $A_{o,t}$ is the best application of the idea in region o at time t and held fixed when computing q:

$$\begin{aligned}
\lambda_t^Z(z_1,\ldots,z_N) &= \sum_{r=1}^{N} \int_0^t \int_0^\infty \left[\min_o \left\{\frac{z_o}{A_{o,t}}\right\}\right]^{-\theta} dF^A_{r,tt^*}(a_1,\ldots,a_N)\ \lambda_{r,t^*} dt^* \\
&= \sum_{r=1}^{N} \int_0^t \int_0^\infty \max_o \left\{A_o^\theta z_o^{-\theta}\right\} dF^A_{r,tt^*}(a_1,\ldots,a_N)\ \lambda_{r,t^*} dt^* \\
&= \sum_{r=1}^{N} \int_0^t \Gamma\left(1-\frac{\theta}{\sigma}\right) \cdot \left[\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \sum_{o=1}^{N} \Omega_{ro,t^*}(t-t^*)\cdot z_o^{-\sigma}\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*} dt^* \\
&= \sum_{r=1}^{N} \int_0^t \left[\sum_{o=1}^{N} \Omega_{ro,t^*}(t-t^*)\cdot z_o^{-\sigma}\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*} dt^*
\end{aligned} \tag{57}$$

where the second equality comes from solving the integral:

$$\int_0^\infty \max_o \left\{A_o^\theta z_o^{-\theta}\right\} dF^A_{r,tt^*}(a_1,\ldots,a_N) = \mathbb{E}\left[\max_o \left\{A_o^\theta Z_o^{-\theta}\right\}\right] \tag{58}$$

via three steps: (i) computing the distribution of the transformed variable $B_o \equiv \frac{a_o}{z_o}$:

$$\mathbb{P}(B_o \le b) = \mathbb{P}\left(\frac{A_o}{z_o} \le b\right) = \mathbb{P}\left(A_o \le b z_o\right) = \exp\left[-\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \Omega_{ro,t^*}(t-t^*)\cdot b^{-\sigma} z_o^{-\sigma}\right],$$

(ii) computing the distribution of the maximum of the transformed variable across regions o, $\max_o B_o$:

$$\mathbb{P}\left(\max_o B_o \le b\right) = \exp\left[-\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \sum_o \Omega_{ro,t^*}(t-t^*)\cdot z_o^{-\sigma}\cdot b^{-\sigma}\right],$$

and (iii) computing the moment $\mathbb{E}\left[\max_o B_o^\theta\right]$ using the known formula for Frechet random variables:

$$\mathbb{E}\left[\max_o B_o^\theta\right] = \mathbb{E}\left[\left(\max_o B_o\right)^\theta\right] = \Gamma\left(1-\frac{\theta}{\sigma}\right)\left[\Gamma\left(1-\frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \sum_o \Omega_{ro,t^*}(t-t^*)\cdot z_o^{-\sigma}\right]^{\frac{\theta}{\sigma}}.$$

Using this Poisson process, the number of ideas with efficiency above z_o in at least one region by time t is distributed Poisson with parameter $\lambda_t^Z(z_1, \ldots, z_N)$:

$$\mathbb{P}\left[I = i\right] = e^{-\lambda_t^Z(z_1,\ldots,z_N)} \frac{\left[\lambda_t^Z(z_1, \ldots, z_N)\right]^i}{i!}.$$

The probability that the i'th most efficient idea has efficiency below z is the probability that at most $i-1$ ideas have efficiency greater than z:

$$F_t^{Z(i)}(z_1, \ldots, z_N) = e^{-\lambda_t^Z(z_1,\ldots,z_N)} \sum_{l=0}^{i-1} \frac{\left[\lambda_t^Z(z_1, \ldots, z_N)\right]^l}{l!}.$$

The joint productivity distribution across all regions at time t is given by the distribution of the most efficient idea:

$$\mathbb{P}[Z_1 \leq z_1, \ldots, Z_N \leq z_n] = F_t^{Z(1)}(z_1, \ldots, z_N) = \exp\left[-\sum_{r=1}^{N} \int_0^t \left[\sum_{o=1}^{N} \Omega_{ro,t^*}(t-t^*) \cdot z_o^{-\sigma}\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*} dt^*\right],$$

which corresponds to equation (13) in the main text. □

C.2 Lemma 2 (Distribution of the Minimum Costs and Trade Shares)

Proof. I first derive the unconditional cost distribution, then the conditional cost distribution in each region, the joint cost distribution, and finally the equilibrium trade shares.

(i) Distribution of the Minimum Cost in each Destination

Each destination sources goods from the cheapest location. Thus, what matters is the number of ideas whose goods can be delivered to region d below cost c by some region o by time t. The expected number of these ideas is the expected number of ideas with efficiency greater than $z_o = \frac{w_{o,t}\tau_{od,t}}{c}$ in at least one region o:

$$\lambda_{d,t}^C(c) = \lambda_t^Z\left(\frac{w_{1,t}\tau_{1d,t}}{c}, \ldots, \frac{w_{N,t}\tau_{Nd,t}}{c}\right) = c^\theta \sum_{r=1}^{N} \int_0^t \left[\sum_{o=1}^{N} \Omega_{ro,t^*}(t-t^*) \cdot (w_{o,t}\tau_{od,t})^{-\sigma}\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*} dt^*. \tag{59}$$

Thus, the number of ideas whose goods can be delivered to region d below cost c by time t – regardless of production and innovation location – is distributed Poisson with parameter $\lambda_{d,t}^C(c)$. Ranking these ideas by their costs in destination d, $C^{(1)} < C^{(2)} < C^{(3)} < \ldots$, the probability that the i'th idea is above cost c is given by:

$$e^{-\lambda_{d,t}^C(c)} \sum_{l=0}^{i-1} \frac{\left[\lambda_{d,t}^C(c)\right]^l}{l!}$$

Thus the probability that the i'th idea is below cost c is given by:

$$G_{d,t}^{(i)}(c) = 1 - e^{-\lambda_{d,t}^{C}(c)} \sum_{l=0}^{i-1} \frac{\left[\lambda_{d,t}^{C}(c)\right]^{l}}{l!} = 1 - \mathbb{P}\left[C_{1d,t}^{(i)} > c, \ldots, C_{Nd,t}^{(i)} > c\right]. \tag{60}$$

In particular, the distribution of the lowest cost is given by:

$$G_{d,t}^{(1)}(c) = 1 - e^{-\lambda_{d,t}^{C}(c)} = 1 - \exp\left\{ - \left[\sum_{r=1}^{N} \int_{0}^{t} \left[\sum_{o=1}^{N} \Omega_{ro,t}(t-t^{*}) \left(w_{o,t}\tau_{od,t}\right)^{-\sigma} \right]^{\frac{\theta}{\sigma}} \lambda_{r,t^{*}} dt^{*} \right] c^{\theta} \right\}.$$

(ii) Equilibrium Trade Shares

Because different production locations may potentially produce goods from the same idea, the number of potential goods produced by each region o is not independent. This correlation across production locations implies that the straightforward way of computing trade shares does not hold, $\pi_{od,t} \neq \frac{\lambda_{od,t}^{C}(c)}{\lambda_{d,t}^{C}(c)}$. Instead, trade shares take on a nested form reflecting this correlation.

Since the number of ideas is independent across research locations, the share of goods purchased by destination d that were produced from ideas discovered in research location r is given by:

$$\phi_{rd,t} = \frac{\lambda_{rd,t}^{C}(c)}{\lambda_{d,t}^{C}(c)} = \frac{\int_{0}^{t} \left[\sum_{o=1}^{N} \Omega_{ro,t^{*}}(t-t^{*}) \cdot \left(w_{o,t}\tau_{od,t}\right)^{-\sigma} \right]^{\frac{\theta}{\sigma}} \lambda_{r,t^{*}} dt^{*}}{\sum_{r'=1}^{N} \int_{0}^{t} \left[\sum_{o=1}^{N} \Omega_{r'o,t^{*}}(t-t^{*}) \cdot \left(w_{o,t}\tau_{od,t}\right)^{-\sigma} \right]^{\frac{\theta}{\sigma}} \lambda_{r',t^{*}} dt^{*}}.$$

For each idea discovered in region r at time t^{*} whose good is eventually sold to region d at time t, its quality q and cost c are fixed and the number of applications in each production location o is independent. Thus, the share of these ideas whose goods are produced in region o is given by the number of applications with applicability above $a = \frac{w_{o,t}\tau_{od,t}}{q \cdot c}$ in region o divided by the total number of applications with applicability above $a = \frac{w_{o',t}\tau_{o'd,t}}{q \cdot c}$ for all regions o':

$$\varphi_{o|rd,t^{*}t} = \frac{\lambda_{o|rd,t}^{A}\left(\frac{w_{o,t}\tau_{od,t}}{q \cdot c}\right)}{\sum_{o'} \lambda_{o'|rd,t}^{A}\left(\frac{w_{o',t}\tau_{o'd,t}}{q \cdot c}\right)} = \frac{\Omega_{ro,t^{*}}(t-t^{*}) \left(w_{o,t}\tau_{od,t}\right)^{-\sigma}}{\sum_{o'=1}^{N} \Omega_{ro',t^{*}}(t-t^{*}) \left(w_{o',t}\tau_{o'd,t}\right)^{-\sigma}}.$$

where the function for the Poisson arrival of applications $\lambda^{A}(a)$ above a comes from equation (10) in the main text. Since this share is independent of q and hence independent across ideas, the share of goods sold in destination d at time t whose ideas were discovered in region r and goods produced in region o (the

trilateral share) is given by:

$$\pi_{rod,t} = \int_0^t \frac{\left[\sum_{o'=1}^{N} \Omega_{ro',t^*}(t-t^*)\cdot\left(w_{o',t}\tau_{o'd,t}\right)^{-\sigma}\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*}}{\sum_{r'=1}^{N}\int_0^t \left[\sum_{o'=1}^{N} \Omega_{r'o',t^*}(t-t^*)\cdot\left(w_{o',t}\tau_{o'd,t}\right)^{-\sigma}\right]^{\frac{\theta}{\sigma}} \lambda_{r',t^*}dt^*} \frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}\tau_{od,t}\right)^{-\sigma}}{\sum_{o'=1}^{N} \Omega_{ro',t^*}(t-t^*)\left(w_{o',t}\tau_{o'd,t}\right)^{-\sigma}} dt^*.$$

The trade shares are obtained by summing these trilateral shares across research locations: $\pi_{od,t} = \sum_r \pi_{rod,t}$. These expressions correspond to equations (18)-(21) in the main text. □

C.3 Lemma 3 (Price Index)

Proof. Under Bertrand competition, consumers regard different varieties (i.e. different idea applications) of a good as perfect substitutes. Since all goods are essential for the production of the local final good, the price charged for each good is its second lowest cost, drawn from the following distribution:

$$G_{d,t}^{(2)}(c) = 1 - e^{-\lambda_{d,t}^{C}(c)}\left[1 + \lambda_{d,t}^{C}(c)\right] \tag{61}$$

where $\lambda_{d,t}^{C}(c)$ comes from equation (59). Notice that this is a special case of equation (60) with $i = 2$. To simplify the subsequent computations, I define:

$$\tilde{\Phi}_{d,t} \equiv \sum_{r=1}^{N}\int_0^t \left[\sum_{o=1}^{N} \Omega_{ro,t^*}(t-t^*)\cdot\left(w_{o,t}\tau_{od,t}\right)^{-\sigma}\right]^{\frac{\theta}{\sigma}} \lambda_{r,t^*}dt^*, \tag{62}$$

so that $\lambda_{d,t}^{C}(c) = \tilde{\Phi}_{d,t}c^{\theta}$ and $G_{d,t}^{(2)}(c) = 1 - e^{-\tilde{\Phi}_{d,t}c^{\theta}}\left[1 + \tilde{\Phi}_{d,t}c^{\theta}\right]$.

Aggregating prices across all varieties yields the price index in each region:

$$P_{d,t} = \exp\int_0^1 \ln p_{d,t}(\nu)d\nu = \exp\int_0^{\infty} \ln c \; dG_{d,t}^{(2)}(c) = \gamma\tilde{\Phi}_{d,t}^{-\frac{1}{\theta}}$$

where γ is the Euler-Mascheroni constant. This corresponds to equation (24) in the main text. □

C.4 Lemma 4 (Markup Distribution)

Proof. Under Bertrand competition, the lowest cost producer of each good claims the entire market for that good, charging the highest markup that deters any competitor from entering. With the Cobb-Douglas aggregation across individual goods to produce the local sectoral final good (equation 16), the price of each good is given by the second lowest cost. Consequently, the markup of each good is the ratio of the lowest cost to the second lowest cost. I now apply the general derivation of the price gap distribution from Eaton and Kortum (2024) in my setting with idea applications, then use it to obtain the ratio of the lowest cost to the second lowest cost.

(i) Distribution of Transformed Costs

To derive the price gap distribution, it is helpful to work with transformed costs $U \equiv \tilde{\Phi} C^{\theta}$ where $\tilde{\Phi}$ is given by equation (62). I first derive the distribution for $U^{(i)}$. Ranking the $U^{(i)}$'s like the $C^{(i)}$'s, the distribution of the lowest transformed cost is the probability that there is no transformed cost below u:

$$\mathbb{P}\left[U^{(1)} \leq u\right] = 1 - e^{-u},$$

which is the unit exponential distribution. Given a lower transformed cost $U^{(i)} = u_i$, the probability that the next lowest transformed cost $U^{(i+1)}$ is less than u is simply the probability that there is no transformed cost between u_i and u:

$$\mathbb{P}\left[U^{(i+1)} \leq u | U^{(i)} = u_i\right] = 1 - e^{-(u-u_i)}. \tag{63}$$

Thus, the distribution of the gap conditional on the lower transformed cost is unit exponential. This implies that the set of ordered transformed costs can be drawn from independent unit exponential distributions V_i where:

$$U^{(1)} = V_1, \qquad U^{(2)} = U^{(1)} + V_2, \qquad U^{(3)} = U^{(2)} + V_3, \qquad \ldots. \tag{64}$$

Since $U^{(i)}$ is the sum of i independent unit exponential distributions, it follows an Erlang distribution with shape parameter i and rate parameter 1. Its probability density function is:

$$f^{(i)}(u) = \frac{u^{i-1} e^{u}}{(i-1)!} \tag{65}$$

and cumulative distribution function is:

$$F^{U(i)}(u) = \mathbb{P}\left[U^{(i)} \leq u\right] = 1 - e^{-u} \sum_{l=0}^{i-1} \frac{u^i}{i!} = \int_0^u \frac{x^{i-1} e^{x}}{(i-1)!}\, dx. \tag{66}$$

Notice that the distribution of transformed costs can also be directly obtained from the cost distribution (equation 60). This derivation provides intuition on why the cost and transformed cost distributions are Erlang.

(ii) Conditional Distribution of Lower Transformed Costs Given Higher Transformed Costs

Using the distribution of transformed costs $U^{(i)}$, I now derive the probability of the lower cost given a higher cost. The joint probability density function of $U^{(i)}$ and $U^{(i+1)}$ is given by:

$$\begin{aligned} f^{(i,i+1)}(u_i, u_{i+1}) &= \frac{d}{du_{i+1}} \left\{ \mathbb{P}\left[U^{(i+1)} \leq u_{i+1} | U^{(i)} = u_i\right] \cdot \mathbb{P}\left[U^{(i)} = u_i\right] \right\} \\ &= \frac{d}{du_{i+1}} \left\{ \left(1 - e^{-(u_{i+1}-u_i)}\right) \frac{u_i^{i-1} e^{u_i}}{(i-1)!} \right\} \\ &= \frac{u_i^{i-1} e^{u_{i+1}}}{(i-1)!} \end{aligned} \tag{67}$$

Thus the conditional distribution of the lower transformed costs (i.e. higher quality draw) is given by:

$$\mathbb{P}\left[U^{(i)} \leq u | U^{(i+1)} = u_{i+1}\right] = \int_0^u \frac{f^{(i,i+1)}(x, u_{i+1})}{f^{(i+1)}(u_{i+1})} dx = \int_0^u \frac{x^{i-1} e^{u_{i+1}}}{(i-1)!} \frac{i!}{u_{i+t}^i e^{u_{i+1}}} dx = \left(\frac{u}{u_{i+1}}\right)^i \tag{68}$$

which is a Pareto distribution with shape parameter i and scale parameter u_{i+1}.

(iii) Markup Distribution under Bertrand Competition

Using this result, the markup distribution conditional on the second lowest cost is given by:

$$\begin{aligned} G^{(2)/(1)}(m) &\equiv \mathbb{P}\left[\frac{C^{(2)}}{C^{(1)}} \leq m | C^{(2)} = c_2\right] = \mathbb{P}\left[\frac{U^{(2)}}{U^{(1)}} \leq m^\theta | U^{(2)} = u_2\right] \\ &= \mathbb{P}\left[U^{(1)} \geq u_2 m^{-\theta} | U^{(2)} = u_2\right] = 1 - \left(\frac{m^{-\theta} u_2}{u_2}\right) = 1 - m^{-\theta}. \end{aligned} \tag{69}$$

Remarkably, this distribution is independent of the second lowest cost c_2. Intuitively, this result arises because the conditional distribution of the higher quality draw is Pareto and hence scale invariant. That this, the ratio of the highest quality to the second highest quality idea is independent of the second highest quality idea.

Note that the markup distribution conditional on the lowest cost is different and given by:

$$\mathbb{P}\left[\frac{C^{(2)}}{C^{(1)}} \leq m | C^{(1)} = c_1\right] = \mathbb{P}\left[U^{(1)} \leq u_1 m^\theta | U^{(1)} = u_1\right] = 1 - \exp\left[-\Phi c_1^\theta \left(m^{1/\theta} - 1\right)\right] \tag{70}$$

In contrast to equation (69), this distribution depends on the lowest cost. Intuitively, a higher value of the higher quality draw implies that the cutoff/second highest quality draw must have been higher. □

C.5 Lemma 5 (Expected Value of an Idea)

Proof. Total profits earned at time s in region d is given by $\Pi_{d,t}^k = \frac{X_{d,t}^k}{1+\theta}$, as described in equation (26). The share of these profits from ideas discovered in region r at time t^* is ϕ_{rd,t^*t}^k. Since the flow rate of ideas in region r at time t^* is λ_{r,t^*}^k, the expected flow of profits at time s in region d of an idea discovered in region r at time t^* is $\frac{\phi_{rd,t^*t}^k}{\lambda_{r,t^*}^k} \frac{X_{d,t}^k}{1+\theta}$. Accounting for changes in the purchasing power in region r over time and discounting future flows yields equation (27). Notice that ϕ_{ld,t^*t}^k conditions on idea cohort and hence is a direct measure on whether the idea remains the lowest cost one in destination d. □

C.6 Lemma 6 (Market Clearing)

Proof. The market clearing condition is slightly more complex relative to standard trade models due to the transfer of profits across regions. To account for these transfers, I distinguish across profits earned from innovation $\overline{\Pi}$, production Π^*, and sales Π in each region. Income in each region $Y_{d,t}$ equals total expenditure

from the region, which comes from two sources: final spending by production workers and profits earned from innovation by firms. Thus the market clearing condition is given by:

$$\begin{aligned} w_{o,t}^{k,G} L_{o,t}^{k,G} + \Pi_{o,t}^{k\ *} &= \sum_d \pi_{od,t}^k \iota^k \left[\sum_k w_{d,t}^{k,G} L_{d,t}^{k,G} + \overline{\Pi_{d,t}^k} \right] \\ \implies \frac{1+\theta}{\theta} w_{o,t}^{k,G} L_{o,t}^{k,G} &= \sum_d \pi_{od,t}^k \iota^k \left[\sum_k \left(w_{d,t}^{k,G} L_{d,t}^{k,G} + \sum_r \varphi_{dr,t}^k \Pi_{r,t}^{k\ *} \right) \right] \\ \implies \frac{1+\theta}{\theta} w_{o,t}^{k,G} L_{o,t}^{k,G} &= \sum_d \pi_{od,t}^k \iota^k \left[\sum_k \left(w_{d,t}^{k,G} L_{d,t}^{k,G} + \sum_r \varphi_{dr,t}^k \frac{1}{\theta} w_{r,t}^{k,G} L_{r,t}^{k,G} \right) \right] \end{aligned} \tag{71}$$

where $\varphi_{dr,t}^k$ are the idea adoption shares and $\pi_{od,t}^k$ are the trade shares. Notice that I eliminate the profit terms in the market clearing condition by exploiting the fact that profits earned from production is a constant multiple of income earned by production workers in the region-sector:

$$w_{o,t}^{k,G} L_{o,t}^{k,G} + \frac{1}{1+\theta} \sum_d \pi_{od,t}^k X_{d,t}^k = \sum_d \pi_{od,t}^k X_{d,t}^k.$$

This enhances the tractability and simplifies the potential quantification of the model. □

C.7 Lemma 7 (Expected Worker Value and Worker Mobility Shares)

Proof. Note that the individual worker mobility problem, as defined by equations 31-32, represents a continuous time extension of Caliendo et al. (2019) (henceforth CDP) and incorporates switching between production and research alongside mobility across regions and sectors. In what follows, I demonstrate how the equilibrium mobility shares, as expressed in equation 34, can be derived under these extensions.

First, transitioning from discrete to continuous time with the possibility of worker movement complicates the distinction between current versus future payoffs. To overcome this challenge, I assume that there is a Poisson arrival process governing when *all* workers can move, with an arrival rate of 1, as described in Assumption 3 in the main text. This setup is a simplified version of Arcidiacono et al. (2016). With a Poisson arrival process of rate 1, the time until the next arrival at t' follows an exponential distribution with rate 1, where the probability density function is given by $e^{-(t'-t)}$. The present value at time t of the payoff at time t' is given by: $V_{t'} e^{-\zeta(t'-t)}$, where ζ is the discount rate. Thus, the expected value at time t of the payoff at time t' is given by:

$$\text{Expected Payoff} = V_{t'} \cdot \int_0^\infty e^{-\zeta(t'-t)} e^{-(t'-t)}\ d(t'-t) = \frac{1}{1+\rho} V_{t'}.$$

This is why the individual worker mobility problem in equation 31 has the discount factor $\frac{1}{1+\rho}$. Note that my formulation is a strict generalization of CDP. Though my discount factor $\frac{1}{1+\rho}$ is isomorphic to β in the migration problem in CDP and $\mathbb{E}_t(t') = t+1$, the actual migration time t' is stochastic and depends on the exact realization of the Poisson process for move arrivals.

Second, I incorporate switching between production and research, where individual-specific idiosyncratic shocks for each market are drawn from a multivariate Gumbel or generalized extreme value distribution with symmetric correlation between production and research across all region-sectors. The functional form of this distribution is provided in equation 32, with the correlation function defined as:

$$\check{F}\left(\left\{\exp\left(-\epsilon_{o,t}^{s,n}\right)\right\}_{o=1,\ldots,N}^{s=\{\text{ICT,non-ICT}\},n=\{G,R\}}\right)=\sum_{o}\sum_{s}\left(\sum_{n}\exp\left(-\epsilon_{o,t}^{s,n}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}, \tag{72}$$

where υ captures the correlation between production and research across region-sectors, and Υ is the scale parameter that adjusts the sensitivity of the expected value in production or research to differences in the deterministic components of value in these activities. Note that this correlation function is homogeneous of degree $\log\Upsilon$ (instead of 1), similar to Ben-Akiva and Francois (1983), but otherwise retains the same properties listed in McFadden (1978).

I now use this correlation function to derive the option value of moving and the resulting mobility shares. Since the option value of moving is defined as:

$$\Phi_{d,t}^{k,h}=\mathbb{E}_{\epsilon}\left[\max_{o,s,n}\left\{\frac{1}{1+\rho}V_{o,t'}^{s,n}-\kappa_{do,t}^{ks,hn}+\epsilon_{o,t}^{s,n}\right\}\right],$$

I first derive the distribution of the maximum of random variables whose joint distribution follows the GEV distribution in equation (32). Let $U_{o,t}^{s,n}=\frac{1}{1+\rho}V_{o,t'}^{s,n}-\kappa_{do,t}^{ks,hn}+\epsilon_{o,t}^{s,n}$ and $U=\max_{o,s,n}U_{o,t}^{s,n}$. Extending Proposition 1 in Choi and Moon (1997), the cumulative distribution function is:

$$\begin{aligned}
&F_U(u)\equiv\mathbb{P}(U\leq u)\\
&=\mathbb{P}(U_1^{\text{ICT},G}\leq u,\ldots,U_N^{\text{non-ICT},R}\leq u)\\
&=\exp\left\{-\check{F}\left(\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{ICT},G}-\kappa_{d1,t}^{k\text{ICT},hG}-u\right],\ldots,\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{non-ICT},R}-\kappa_{dN,t}^{k\text{non-ICT},hR}-u\right]\right)\right\}\\
&=\exp\left\{-\exp(-\Upsilon u)\check{F}\left(\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{ICT},G}-\kappa_{d1,t}^{k\text{ICT},hG}\right],\ldots,\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{non-ICT},R}-\kappa_{dN,t}^{k\text{non-ICT},hR}\right]\right)\right\}\\
&=\exp\left\{-\exp\left(-\Upsilon\left[u-\frac{1}{\Upsilon}\log\left[\check{F}\left(\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{ICT},G}-\kappa_{d1,t}^{k\text{ICT},hG}\right],\ldots,\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{non-ICT},R}-\kappa_{dN,t}^{k\text{non-ICT},hR}\right]\right)\right]\right]\right)\right\}
\end{aligned}$$

where the second last equality comes from the correlation function being homogeneous of degree $\log\Upsilon$. Thus, the distribution of the maximum is a Gumbel distribution with scale parameter Υ and location parameter $\frac{1}{\Upsilon}\log\left[\check{F}\left(\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{ICT},G}-\kappa_{d1,t}^{k\text{ICT},hG}\right],\ldots,\exp\left[\frac{1}{1+\rho}V_{1,t'}^{\text{non-ICT},R}-\kappa_{dN,t}^{k\text{non-ICT},hR}\right]\right)\right]$. The mean of this distribution is $\frac{1}{\Upsilon}\log\check{F}+\Upsilon\gamma$, where γ is Euler's constant. Thus the option value of moving – also known as the inclusive value of the entire choice set in the discrete choice literature – is given by:

$$\Phi_{d,t}^{k,h}=\frac{1}{\Upsilon}\log\left[\sum_{o,s}\left(\sum_{n}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n}-\kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right], \tag{73}$$

where the term $\Upsilon\gamma$ is constant over time and markets and hence can be normalized to 0. Substituting the

option value of moving into the individual worker migration problem in equation 31 yields the expected worker value in equation 33.

Mobility shares are given by the probability of choosing a specific market (o, s, n) to move to:

$$\begin{aligned}\mu_{do,t}^{ks,hn} &= \frac{\partial \Phi_{d,t}^{k,h}}{\partial \left(\frac{1}{1+\rho} V_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)} \\ &= \underbrace{\frac{\exp\left(\frac{1}{1+\zeta} V_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'} \exp\left(\frac{1}{1+\zeta} V_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}}}_{\text{switching between production and research}} \cdot \underbrace{\frac{\left[\sum_{n'} \exp\left(\frac{1}{1+\zeta} V_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{o'} \sum_{s'} \left[\sum_{n'} \exp\left(\frac{1}{1+\zeta} V_{o',t'}^{s',n'} - \kappa_{do',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}}_{\text{migration across regions and sectors}}\end{aligned}$$

where the first equality comes from McFadden (1978). Note that given the correlation function, the individual worker mobility problem can also be recast as a nested discrete choice problem: each worker first chooses which region-sector to move to and then whether to supply their labor for research or production in that region-sector. In this formulation, we can interpret the second term as the probability of choosing region o and sector s among all region-sector alternatives and the first term as the conditional probability of choosing production or research given the region-sector choice. The conditional option value of choosing between production and research given the choice of region o and sector s is:

$$\Phi_{do,t}^{ks,h} = \frac{\upsilon}{\Upsilon} \log\left[\sum_{n} \exp\left(\frac{1}{1+\zeta} V_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right].$$

The unconditional option value across all regions, sectors, and occupations – as shown in equation 73 – is related to this conditional option value as follows:

$$\Phi_{d,t}^{k,h} = \frac{1}{\upsilon} \log\left[\sum_{o,s} \exp\left(\Phi_{do,t}^{ks,h}\right)^{\upsilon}\right]. \tag{74}$$

Notice that when $\upsilon = 1$ in the correlation function, draws of ϵ are independent across markets, and the worker mobility shares and option value collapses to expressions of the form in Caliendo et al. (2019). By leveraging the correlation function and techniques from the foundational discrete choice literature, this proof *clarifies and generalizes dynamic migration in general equilibrium to a setting where preference shocks are correlated across markets*, i.e. drawn from a generalized extreme value distribution with any arbitrary correlation function $\check{F}$ that satisfies the properties listed in McFadden (1978) and μ-homogeneity in Ben-Akiva and Francois (1983). □

C.8 Proposition 1 (Spatial and Sectoral Direction of Innovation)

The ratio of inventor real wages across regions in the same sector follows directly from the text. Additionally, the sectoral direction of innovation is governed by:

$$\frac{\omega^{k,R}_{r,t^*}}{\omega^{k',R}_{r,t^*}} = \underbrace{\frac{A^k_{t^*}}{A^{k'}_{t^*}}}_{\substack{\text{differences in}\\ \text{fundamental}\\ \text{research}\\ \text{productivity}\\ \text{across sectors}}} \cdot \underbrace{\frac{T^k_{r,t^*}}{T^{k'}_{r,t^*}}}_{\substack{\text{relative}\\ \text{technology}\\ \text{levels}}} \cdot \underbrace{\left(\frac{L^{k,G}_{r,t^*}}{L^{k',G}_{r,t^*}}\right)^{\chi}}_{\substack{\text{benefits of}\\ \text{colocation of}\\ \text{innovation and}\\ \text{production}}} \cdot \underbrace{\frac{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\phi^k_{rd,t^*t}}{\lambda^k_{r,t^*}} \frac{X^k_{d,t}}{1+\theta} \frac{1}{P_{r,t}} dt}{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\phi^{k'}_{rd,t^*t}}{\lambda^{k'}_{r,t^*}} \frac{X^{k'}_{d,t}}{1+\theta} \frac{1}{P_{r,t}} dt}}_{\text{expected market potential of an idea}}. \tag{75}$$

The first term captures the one-time increase in ICT research productivity at a national level (the first exogenous component of the ICT shock), the second term captures relative technology levels across sectors, the third term anchors the field of innovative activity to the sectoral composition of production in the region, and the fourth term captures differences in the expected market potential of ideas across sectors.

C.9 Proposition 2 (Balanced Growth Path)

Proof. **Part (i)**: On the balanced growth path, technology levels in all regions and sectors grow at the same rate $g = \frac{\dot{T}^k_r(t)}{T^k_r} \forall r, k$, the exogenous bilateral diffusion lags δ_{ro} are constant over time and the innovation rate is given by:

$$\lambda^k_r(t) = \gamma^k_r T^k_r(t) \implies \dot{\lambda}^k_r(t) = \gamma^k_r \dot{T}^k_r(t). \tag{76}$$

From equation 15 we know that the technology level at each time t is given by:

$$T^k_{o,t} = \sum_{r=1}^{N} T^k_{ro,t} = \sum_{r=1}^{N} \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} \cdot \lambda^k_r(t^*) dt^*.$$

Taking the derivative w.r.t. t yields:

$$\dot{T}^k_{o,t} = \sum_{r=1}^{N} \int_{-\infty}^{t} \frac{d\Omega_{ro}(t-t^*)^{1-\rho}}{dt} \cdot \lambda^k_r(t^*) dt^* + \Omega_{ro}(0)^{1-\rho} \lambda^k_r(t).$$

Now using integration by parts with $u = \lambda^k_r(t^*)$ and $dv = \frac{d\Omega_{ro}(t-t^*)^{1-\rho}}{dt}$ yields:

$$\begin{aligned}
\dot{T}^k_{o,t} &= \sum_{r=1}^{N} \left[-\lambda^k_r(t^*)\Omega_{ro}(t-t^*)^{1-\rho}\right]_{t^*=-\infty}^{t^*=t} + \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} \dot{\lambda}^k_r(t^*) dt^* + \Omega_{ro}(0)^{1-\rho}\lambda_l(t) \\
&= \sum_{r=1}^{N} \lim_{t^* \to -\infty} \lambda^k_r(t^*)\Omega_{ro}(t-t^*)^{1-\rho} + \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} \gamma^k_r \dot{T}_r(t^*) dt^* \\
&= \sum_{r=1}^{N} \gamma^k_r \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} \dot{T}^k_r(t^*) dt^*.
\end{aligned}$$

Thus we have that:

$$\begin{aligned}\dot{T}^k_{o,t} &= \sum_{r=1}^{N} \gamma^k_r \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} \frac{\dot{T}^k_r(t^*)}{T^k_l(t^*)} T^k_r(t^*) dt^* \\ &= g \sum_{r=1}^{N} \gamma^k_r \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} T^k_r(t) e^{-g(t-t^*)} dt^* \\ &= \sum_{r} \gamma^k_r T^k_{r,t} \int_{-\infty}^{t} g e^{-g(t-t^*)} \left[1 - e^{-\delta_{ro}(t-t^*)}\right]^{1-\rho} dt^* \\ &= \sum_{r} \gamma^k_r T^k_{r,t} \int_{0}^{\infty} g e^{-ga} \left[1 - e^{-\delta_{ro}(a)}\right]^{1-\rho} da\end{aligned} \tag{77}$$

where $e^{-\delta_{rr}(t-t^*)} \equiv 0$ and the last equality follows from a change of variable $a = t - t^*$. Note that $\int_0^\infty g e^{-ga} \left[1 - e^{-\delta_{ro}(a)}\right]^{1-\rho} da$ is a constant.

Now building on Eaton and Kortum (2024), in matrix form we have:

$$g\mathbf{T}^k = \mathbf{\Delta}^k(g)\mathbf{T}^k \tag{78}$$

where $\mathbf{T}^k$ is an $N \times 1$ vector with representative element T^k_r and $\mathbf{\Delta}^k(g)$ is an $N \times N$ matrix with representative element:

$$\Delta^k_{ro}(g) = \gamma^k_r \int_0^\infty g e^{-ga} \left[1 - e^{-\delta_{ro}(a)}\right]^{1-\rho} da.$$

The aggregate growth rate is the Perron-Frobenius root of equation 78 with relative technology levels $\mathbf{T}$ corresponding to the Perron-Frobenius eigenvector that is defined up to a scalar multiple. Thus, any arbitrary set of exogenous diffusion speeds δ_{ro} and endogenous innovation rates γ^k_r delivers a balanced growth path with parallel growth at rate g with level differences in technology T across regions.

Part (ii): I now solve for the remaining variables in the economy on the BGP. Suppose that the distribution of workers across regions, sectors, and occupations are constant. Since the relative technology levels are also constant on the BGP, the trade shares given by equation 18 and the market condition given by equation 29 imply that relative wages across regions and sectors in production is constant. In particular, trade shares on the BGP are given by:

$$\pi^k_{od} = \sum_{r=1}^{N} \int_0^\infty \frac{\Omega_{ro}(a) \left(w^k_o \tau^k_{od}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N} \Omega_{ro'}(a) \left(w^k_{o'} \tau^k_{o'd}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\gamma^k_r T^k_r e^{-g^k \cdot a} \left[\sum_{o'=1}^{N} \Omega_{ro'}(a) \left(w^k_{o'} \tau^k_{o'd}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N} \gamma^k_{r'} T^k_{r'} \int_0^\infty \left[\sum_{o'=1}^{N} \Omega_{r'o'}(a') \left(w^k_{o'} \tau^k_{o'd}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} e^{-g^k \cdot a'} da'} da. \tag{79}$$

where T^k_r are the relative technology levels determined by equation 78.

On the BGP, the price index in each region and sector is given by:

$$P_{d,t}^{k\ BGP} = \Gamma \left[\sum_{r'=1}^{N} \gamma_{r'}^{k} T_{r',t}^{k} \int_{0}^{\infty} e^{-ga} \left[\sum_{o'=1}^{N} \Omega_{r'o'}(a) \left(w_{o'}^{k} \tau_{o'd}^{k} \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho} da \right]^{-\frac{1}{\theta}} \tag{80}$$

Differentiating this expression w.r.t time, the growth rate of the sectoral price index is:

$$g_{P^k} = \frac{\dot{P_{d,t}^{k}}^{BGP}}{P_{d,t}^{k\ BGP}} = -\frac{1}{\theta} \sum_{r'} \check{\pi}_{r'd} \frac{\dot{T_{r',t}^{k}}}{T_{r',t}^{k}} = -\frac{g^k}{\theta} \tag{81}$$

Since preferences are Cobb-Douglas across local sectoral final goods for each region, the aggregate price index is:

$$P_{d,t} = \check{\Gamma} \prod_{k} \left(P_{d,t}^{k} \right)^{\iota^k} \tag{82}$$

where $\check{\Gamma}$ is a constant. The growth rate of the aggregate price index on the BGP is:

$$g_p = -\frac{1}{\theta} \sum_{k} \iota^k g^k \tag{83}$$

On the BGP, the expected value of an idea is:

$$\begin{aligned} \check{V}_{r,t^*}^{k} &= \frac{P_{r,t^*}}{1+\theta} \int e^{-\zeta(t-t^*)} \sum_{d} \frac{\phi_{rd,t^*t}}{\gamma_r^k T_{r,t^*}^k} \frac{X_{d,t}}{P_{r,t}} dt \\ &= \frac{\sum_d \phi_{rd} X_d}{1+\theta} \frac{1}{\gamma_r^k T_{r,t^*}^k} \int e^{-\zeta(t-t^*)} e^{-g_P(t-t^*)} dt \\ &= \frac{\sum_d \phi_{rd} X_d}{1+\theta} \frac{1}{\gamma_r^k T_{r,t^*}^k} \frac{1}{\zeta - g_P/\theta}. \end{aligned} \tag{84}$$

Thus $\check{V}_{r,t^*}^{k}$ is falling at rate g^k on the BGP while inventor wages are constant and given by:

$$w_r^{k,R} = \frac{\sum_d \check{\pi}_{rd} X_d}{1+\theta} \frac{1}{L_r^{k,R}} \frac{1}{\zeta - g_P/\theta}. \tag{85}$$

I now solve for the growth rate of worker expected value. Let $\exp\left(V_{d,t}^{k,h}\right) = \exp\left(\widetilde{V}_{d}^{k,h}\right) e^{g_V t}$ and $P_{d,t} = \widetilde{P}_d e^{g_P t}$, where $\widetilde{V}_{d}^{k,h}$ and $\widetilde{P}_d$ are the detrended value and price respectively. Thus given production worker

wages, inventor wages and local aggregate prices, worker expected value is given by:

$$\begin{aligned} V_{d,t}^{k,h} &= \widetilde{V}_d^{k,h} + g_V t \\ &= \log\left(\frac{w_d^{k,h}}{\widetilde{P}_d}\right) - g_P t + \frac{1}{\Upsilon}\log\left[\sum_o \sum_s \left(\sum_n \exp\left(\frac{1}{1+\zeta}\left[\widetilde{V}_o^{s,n} + g_V t'\right] - \kappa_{do}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right] \\ &= \log\left(\frac{w_d^{k,h}}{\widetilde{P}_d}\right) - g_P t + \frac{1}{1+\zeta} g_V t' + \frac{1}{\Upsilon}\log\left[\sum_o \sum_s \left(\sum_n \exp\left(\frac{1}{1+\zeta}\widetilde{V}_o^{s,n} - \kappa_{do}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right] \end{aligned} \tag{86}$$

On the BGP, the growth rate must be the same on both sides of the equation. Thus, the growth rate of expected value is given by:

$$\begin{aligned} g_v &= -g_p + \frac{1}{1+\zeta} g_v \\ \implies g_v &= -\frac{1+\zeta}{\zeta} g_p = \frac{1+\zeta}{\zeta}\frac{1}{\theta}\sum_k \iota^k g^k \end{aligned} \tag{87}$$

where the first equality comes from $\frac{1}{1+\zeta} g_v t' = \frac{1}{1+\zeta} g_v t + \frac{1}{1+\zeta} g_v$ in equation 86 since $\mathbb{E}_t(t'-t) = 1$, because the Poisson arrival rate of move possibilities is 1. The detrended expected value of workers is given by:

$$\widetilde{V}_d^{k,h} = \log\left(\frac{w_d^{k,h}}{\widetilde{P}_d}\right) + \frac{1}{1+\zeta} g_v + \frac{1}{\Upsilon}\log\left[\sum_o \sum_s \left(\sum_n \exp\left(\frac{1}{1+\zeta}\widetilde{V}_o^{s,n} - \kappa_{do}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right]. \tag{88}$$

Substituting the decomposition of expected worker value in equation 86 into equation 34 in the main text, worker mobility shares are alternately given by:

$$\mu_{do}^{ks,hn} = \frac{\exp\left(\frac{1}{1+\zeta}\widetilde{V}_o^{s,n} - \kappa_{do}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'} \exp\left(\frac{1}{1+\zeta}\widetilde{V}_o^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}} \cdot \frac{\left[\sum_{n'} \exp\left(\frac{1}{1+\zeta}\widetilde{V}_o^{s,n'} - \kappa_{do}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{o'}\sum_{s'}\left[\sum_{n'} \exp\left(\frac{1}{1+\zeta}\widetilde{V}_{o'}^{s',n'} - \kappa_{do'}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}} \tag{89}$$

and are constant on the BGP. Hence, from equation 35, the distribution of workers across regions, sectors, and occupations are constant on the BGP.

Thus the balanced growth path of the economy is obtained, where workers, wages, migration shares, trade shares and innovation rates are constant, technology in each sector k and region is growing at rate g^k determined from equation 78, prices in each region is growing at rate $g_p = -\frac{1}{\theta}\sum_k \iota^k g^k$ and expected worker value is growing at rate $g_v = \frac{1+\zeta}{\zeta}\frac{1}{\theta}\sum_k \iota^k g^k$. □

Special Cases:

(i) When $\theta = \sigma$ (such that $\rho = 0$), we have the case of exponential diffusion where idea applicabilities do

not matter (**case NA** for no applicabilities). Equation (77) collapses to:

$$\begin{aligned}\dot{T}_{o,t}^{k} &= \sum_{r} \gamma_{r}^{k} T_{r,t}^{k} \int_{0}^{\infty} g e^{-ga} \left[1 - e^{-\delta_{ro}(a)}\right] da \\ &= \sum_{r} g \gamma_{r}^{k} T_{r,t}^{k} \left(\frac{1}{g} + \frac{1}{g + \delta_{ro}}\right) \\ &= \sum_{r} \left(\frac{\delta_{ro}}{g + \delta_{ro}}\right) \gamma_{r}^{k} T_{r,t}^{k}\end{aligned}$$

yielding equation (49) in Eaton and Kortum (2024). The trade shares and price index collapse to the canonical Eaton and Kortum (2002) expressions:

$$\pi_{od}^{NA} = \frac{T_{o,t} \left(w_o \tau_{od}\right)^{-\theta}}{\sum_{o'=1}^{N} T_{o',t} \left(w_{o'} \tau_{o'd}\right)^{-\theta}}, \quad P_{d,t}^{NA} = \Gamma \left[\sum_{o'=1}^{N} T_{o',t} \left(w_{o'} \tau_{o'd}\right)^{-\theta}\right]^{-\frac{1}{\theta}} \tag{90}$$

and idea diffusion shares are given by:

$$\varphi_{ro}^{NA} = \sum_{d=1}^{N} \frac{T_{ro,t} \left(w_o \tau_{od}\right)^{-\theta}}{\sum_{o'=1}^{N} T_{o',t}^{k} \left(w_{o'} \tau_{o'd}\right)^{-\theta}}. \tag{91}$$

All the other variables and growth rates remain the same as the full model.

(ii) When $\Omega_{ro,t^*}(t - t^*) = \delta_{ro,t}$ for $r \neq o$ and $t \geq t^*$, we have the case of instantaneous diffusion with idea applicabilities (**case ID**) as in Xiang (2023), equation (77) collapses to:

$$\begin{aligned}\dot{T}_{o,t}^{k} &= \sum_{r} \gamma_{r}^{k} T_{r,t}^{k} \int_{0}^{\infty} g e^{-ga} \delta_{ro}^{1-\rho} da \\ &= \sum_{r} \delta_{ro}^{1-\rho} \gamma_{r}^{k} T_{r,t}^{k}.\end{aligned}$$

The trade shares and price index collapse to the expressions in Ramondo and Rodríguez-Clare (2013):

$$\pi_{od}^{ID} = \sum_{r=1}^{N} \frac{\left(T_{ro,t}\right)^{\frac{1}{1-\rho}} \left(w_o \tau_{od}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N} \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o'} \tau_{o'd}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N} \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o'} \tau_{o'd}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N} \left[\sum_{o'=1}^{N} \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o'} \tau_{o'd}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{92}$$

$$P_{d,t}^{ID} = \Gamma \left[\sum_{r'=1}^{N} \left[\sum_{o'=1}^{N} \left(T_{r'o',t}\right)^{\frac{1}{1-\rho}} \left(w_{o'} \tau_{o'd}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\right]^{-\frac{1}{\theta}} \tag{93}$$

and the idea diffusion shares are:

$$\varphi_{od,t}^{ID} = \frac{\left(T_{ro,t}\right)^{\frac{1}{1-\rho}} \left(w_o \tau_{od}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N} \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o'} \tau_{o'd}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N} \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o'} \tau_{o'd}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{l'=1}^{N} \left[\sum_{o'=1}^{N} \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o'} \tau_{o'd}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{94}$$

All the other variables and growth rates remain the same as the full model.

C.10 Proposition 3 (Transition Path)

Proof. I first characterize the transition path with Assumption 3, then with Assumption 4 in order to apply dynamic hat algebra and map my model to data on innovation levels, trade shares, migration shares, and firm establishment networks.

(i) Transition Path with Continuous Innovation, Production, and Consumption; and Migration at Discrete Moments in Time (Assumption 3)

The transition path towards balanced growth is characterized by a trajectory of detrended expected values given by equation 88 and worker mobility shares given by equation 89 but with all variables varying over time, alongside the evolution of worker population given by equation 35:

$$\widetilde{V}_{d,t}^{k,h} = \int_t^{t'} \log\left(\frac{w_{d,\check{t}}^{k,h}}{\widetilde{P}_{d,\check{t}}}\right) d\check{t} + \frac{1}{1+\zeta} g_v + \frac{1}{\Upsilon} \log\left[\sum_o \sum_s \left(\sum_n \exp\left(\frac{1}{1+\zeta} \widetilde{V}_{o,t'}^{s,n} - \kappa_{do}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right]$$

$$\mu_{do,t}^{ks,hn} = \frac{\exp\left(\frac{1}{1+\zeta} \widetilde{V}_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'} \exp\left(\frac{1}{1+\zeta} \widetilde{V}_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}} \cdot \frac{\left[\sum_{n'} \exp\left(\frac{1}{1+\zeta} \widetilde{V}_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{o'} \sum_{s'} \left[\sum_{n'} \exp\left(\frac{1}{1+\zeta} \widetilde{V}_{o',t'}^{s',n'} - \kappa_{do',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}$$

$$L_{o,t'}^{s,n} = \sum_h \sum_k \sum_d \mu_{do,t'}^{ks,hn} L_{d,t}^{k,h},$$

where at each time t the innovation and technology levels are given by equations 36 and 15 respectively:

$$\lambda_{r,t}^k = A_{t^*}^k A_{r,t^*} \left(L_{r,t^*}^{k,G}\right)^{\chi} \left(L_{r,t^*}^R\right)^{\alpha} L_{r,t^*}^{k,R} T_{r,t^*}^k$$

$$T_{o,t}^k = \sum_{r=1}^{N} \int_{-\infty}^{t} \Omega_{ro,t^*} (t - t^*)^{1-\rho} \cdot \lambda_{r,t^*}^k dt^*,$$

the trade shares, idea adoption shares, price index, and market clearing condition are given by equations 18,

22, 24, and 29 respectively:

$$\pi_{od,t}^{k}=\sum_{r=1}^{N}\int_{-\infty}^{t}\frac{\Omega_{ro,t^{*}}(t-t^{*})\left(w_{o,t}^{k,G}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^{*}}(t-t^{*})\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^{*}}(t-t^{*})\left(w_{r,t}^{k,G}\tau_{ro',t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^{*}}^{k}}{\sum_{r'}\int_{-\infty}^{t}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^{k}d\check{t}}dt^{*}$$

$$\varphi_{ro,t}^{k}=\sum_{d=1}^{N}\int_{-\infty}^{t}\frac{\Omega_{ro,t^{*}}(t-t^{*})\left(w_{o,t}^{k,G}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^{*}}(t-t^{*})\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^{*}}(t-t^{*})\left(w_{r,t}^{k,G}\tau_{ro',t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^{*}}^{k}}{\sum_{r'}\int_{-\infty}^{t}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^{k}d\check{t}}dt^{*}$$

$$P_{d,t}^{k}=\gamma\left[\sum_{r'=1}^{N}\int_{-\infty}^{t}\left[\sum_{o'=1}^{N}\Omega_{r'o',t^{*}}(t-t^{*})\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',t^{*}}^{k}dt^{*}\right]^{-\frac{1}{\theta}}$$

$$\frac{1+\theta}{\theta}w_{o,t}^{k,G}L_{o,t}^{k,G}=\sum_{d}\pi_{od,t}^{k}\iota^{k}\left[\sum_{k}\left(w_{d,t}^{k,G}L_{d,t}^{k,G}+\sum_{r}\varphi_{dr,t}^{k}\frac{1+\theta}{\theta}w_{r,t}^{k,G}L_{r,t}^{k,G}\right)\right],$$

the wages of inventors are given by equation (28), the returns to innovation by equation (27), total expenditures by equation (49), and idea market shares by equation (50):

$$w_{r,t}^{k,R}=\frac{\check{V}_{r,t}^{k}\lambda_{r,t}^{k}}{L_{r,t}^{k,R}}$$

$$\check{V}_{r,t}^{k}=\int_{t}^{\infty}e^{-\zeta(t'-t)}\sum_{d=1}^{N}\frac{X_{d,t'}^{k}}{1+\theta}\cdot\frac{P_{r,t}}{P_{r,t'}}\cdot\frac{\check{\pi}_{ld}^{k}(t,t')}{\lambda_{l,t}^{k,\check{R}}}\;dt'$$

$$X_{d,t'}^{k}=\iota^{k}\left[\sum_{k}\left(w_{d,t'}^{k,G}L_{d,t'}^{k,G}+\sum_{l}\varphi_{dr,t'}^{k}\frac{1+\theta}{\theta}w_{r,t'}^{k,G}L_{r,t'}^{k,G}\right)\right]$$

$$\phi_{rd,tt'}^{k}=\frac{\left[\sum_{o'}\Omega_{lo',t}(t'-t)\left(w_{l',t'}^{k}\tau_{lo',t'}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{l,t}^{k,\check{R}}}{\sum_{l'}\int_{-\infty}^{t'}\left[\sum_{o'}\Omega_{l'o',\check{t}}(t'-\check{t})\left(w_{o',t'}^{k}\tau_{o'd,t'}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{l',\check{t}}^{k,\check{R}}d\check{t}}.$$

(ii) Transition Path with Innovation, Production, Consumption, and Migration Occuring Simultaneously at Discrete Moments in Time (Assumption 4)

I now apply Assumption 4 to characterize the transition path in time changes. Let $\mathcal{T}$ be the set of times where innovation, production, consumption, and migration all occur simultaneously, $\dot{x}_t=\frac{x_{t'}}{x_t}$ for any variable x, and $u=\exp(V)$. Then the detrended equations for the migration equilibrium (i.e. equations 88, 89, 35)

can be expressed in time changes:

$$\log\left(\widehat{u}_{d,t}^{k,h}\right) = \log\left(\frac{\widehat{w}_{d,t}^{k,h}}{\widehat{P}_{d,t}}\right) + \frac{1}{\Upsilon}\log\left[\sum_{n}\left[\sum_{s}\sum_{o}\mu_{do,t'}^{ks,hn}\left(\widehat{u}_{o,t'}^{s,n}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}\right] \tag{95}$$

$$\mu_{od,t'}^{ks,hn} = \frac{\mu_{od,t}^{ks,hn}\left(\widehat{u}_{d,t'}^{s,n}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'}\mu_{od,t}^{ks,hn'}\left(\widehat{u}_{d,t'}^{s,n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}} \cdot \frac{\left[\sum_{n'}\mu_{od,t}^{ks,hn'}\left(\widehat{u}_{d,t'}^{s,n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{n'}\left[\sum_{d'}\sum_{s'}\mu_{od',t}^{ks',hn'}\left(\widehat{u}_{d',t'}^{s',n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}} \tag{96}$$

$$L_{d,t'}^{k,h} = \sum_{n}\sum_{s}\sum_{o}\mu_{od,t}^{ks,hn}L_{o,t}^{s,n}. \tag{97}$$

Technology levels is now given by:

$$T_{o,t}^{k} = \sum_{r=1}^{N}\sum_{t^*\in\mathcal{T}_{-\infty}^{t}}\Omega_{ro,t^*}(t-t^*)^{1-\rho}\cdot\lambda_{r,t^*}^{k}, \tag{98}$$

where the diffusion term $\Omega_{ro,t^*}(t-t^*)$ remains unchanged, while innovation levels remain unchanged from equation (36). The trade shares, idea adoption shares, and price index is now given by:

$$\pi_{od,t}^{k} = \sum_{r=1}^{N}\sum_{t^*\in\mathcal{T}}\frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}^{k}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^*}^{k}}{\sum_{r'}\sum_{t^*\in\mathcal{T}}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^{k}} \tag{99}$$

$$\varphi_{ro,t}^{k} = \sum_{d=1}^{N}\sum_{t^*\in\mathcal{T}}\frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}^{k}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^*}^{k}}{\sum_{r'}\sum_{t^*\in\mathcal{T}}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^{k}} \tag{100}$$

$$P_{d,t}^{k} = \Gamma\left[\sum_{r'=1}^{N}\sum_{t^*\in\mathcal{T}}\left[\sum_{o'=1}^{N}\Omega_{r'o',t^*}(t-t^*)\left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',t^*}^{k}\right]^{-\frac{1}{\theta}}, \tag{101}$$

where $\mathcal{T}\equiv\mathcal{T}_{-\infty}^{t}$ are the set of times from $-\infty$ to t where innovation, production, consumption, and migration occurs. The market clearing condition remains unchanged from equation (29). Inventor wages are given by equation (28), total expenditures by equation (49), and the expected value of an idea and the idea market shares are now given by:

$$\check{V}_{r,t}^{k} = \sum_{t'\in\mathcal{T}_{t}^{\infty}}\left(\frac{1}{1+\zeta}\right)^{t'-t}\sum_{d=1}^{N}\frac{X_{d,t'}^{k}}{1+\theta}\cdot\frac{P_{r,t}}{P_{r,t'}}\cdot\frac{\phi_{rd,tt'}^{k}}{\lambda_{r,t}^{k}} \tag{102}$$

$$\phi_{rd,tt'}^{k} = \frac{\left[\sum_{o'}\Omega_{ro',t}(t'-t)\left(w_{r',t'}^{k}\tau_{ro',t'}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t}^{k}}{\sum_{r'}\sum_{\check{t}\in\mathcal{T}_{-\infty}^{t'}}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t'-\check{t})\left(w_{o',t'}^{k}\tau_{o'd,t'}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^{k}}. \tag{103}$$

□

Special Cases:

When idea applicabilities are not relevant (**case NA**), only changes in trade costs – as opposed to levels – along with the other fundamentals are required to simulate the transition path. This is because the trade shares, idea diffusion shares, and price index depend only on the contemporaneous technology stock rather than the trajectory of innovation in all past periods, and are given by equations 90 and 91:

$$\pi_{od,t}^{k,NA} = \frac{T_{o,t}^k \left(w_{o,t}^k \tau_{od,t}^k\right)^{-\theta}}{\sum_{o'=1}^{N} T_{o',t}^k \left(w_{o',t}^k \tau_{o'd,t}^k\right)^{-\theta}},\ \varphi_{ro,t}^{k,NA} = \sum_{d=1}^{N} \frac{T_{ro,t}^k \left(w_{o,t}^k \tau_{od,t}^k\right)^{-\theta}}{\sum_{o'=1}^{N} T_{o',t}^k \left(w_{o',t}^k \tau_{o'd,t}^k\right)^{-\theta}},\ P_{d,t}^{k,NA} = \Gamma \left[\sum_{o'=1}^{N} T_{o',t}^k \left(w_{o',t}^k \tau_{o'd,t}^k\right)^{-\theta}\right]^{-\frac{1}{\theta}}.$$

Thus, the production side of the economy can also be expressed using dynamic hat algebra. Changes in trade shares, idea diffusion shares and the price index are given by:

$$\widehat{\pi}_{od,t}^{k,NA} = \frac{\widehat{T}_{o,t}^k \left(\widehat{w}_{o,t}^k \widehat{\tau}_{od,t}^k\right)^{-\theta}}{\sum_{o'=1}^{N} \widehat{T}_{o',t}^k \left(\widehat{w}_{o',t}^k \widehat{\tau}_{o'd,t}^k\right)^{-\theta}} \tag{104}$$

$$\widehat{\varphi}_{ro,t}^{k,NA} = \sum_{d=1}^{N} \frac{\widehat{T}_{ro,t}^k \left(\widehat{w}_{o,t}^k \widehat{\tau}_{od,t}^k\right)^{-\theta}}{\sum_{o'=1}^{N} \widehat{T}_{o',t}^k \left(\widehat{w}_{o',t}^k \widehat{\tau}_{o'd,t}^k\right)^{-\theta}} \tag{105}$$

$$\widehat{P}_{d,t}^{k,NA} = \left[\sum_{o'=1}^{N} \widehat{T}_{o',t}^k \left(\widehat{w}_{o',t}^k \widehat{\tau}_{o'd,t}^k\right)^{-\theta}\right]^{-\frac{1}{\theta}} \tag{106}$$

and the market clearing condition remains unchanged from equation 29. Innovation levels remain unchanged from equation 36 and changes in technology levels are given by:

$$\begin{aligned} T_{o,t'}^k - T_{o,t}^k &= \sum_r \delta_{ro,t} \left(\Lambda_{l,t}^k - T_{lo,t}^k\right) \\ \implies \widehat{T}_{o,t'}^k &= 1 - \delta_{ro,t} + \sum_l \frac{\delta_{lo,t} \Lambda_{l,t}^k}{T_{o,t}^k} \end{aligned} \tag{107}$$

where $\Lambda_{r,t}^k = \sum_{t^* \in \mathcal{T}} \lambda_{r,t^*}^k$ is the stock of innovations produced in region r at time t. This recursive formulation of exponential idea diffusion comes from Eaton and Kortum (2024). Since changes in technology are a function of the previous technology level, data on initial technology levels are still required to solve the transition path quantitatively.

Similarly, when there is instantaneous diffusion (**case ID**), only changes in trade costs – as opposed to levels – along with the other fundamentals are required to simulate the transition path. Trade shares, idea

diffusion shares, and price index are given by equations 92 -94:

$$\pi_{od,t}^{k,ID} = \sum_{r=1}^{N} \frac{\left(T_{ro,t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o,t}^{k}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N}\left(T_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N}\left(T_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N}\left[\sum_{o'=1}^{N}\left(T_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}$$

$$P_{d,t}^{k,ID} = \Gamma\left[\sum_{r'=1}^{N}\left[\sum_{o'=1}^{N}\left(T_{r'o',t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\right]^{-\frac{1}{\theta}}$$

$$\varphi_{od,t}^{k,ID} = \frac{\left(T_{ro,t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o,t}^{k}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N}\left(T_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N}\left(T_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N}\left[\sum_{o'=1}^{N}\left(T_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}^{k}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}$$

Thus changes in trade shares, idea diffusion shares and the price index are given by:

$$\widehat{\pi}_{od,t}^{k,ID} = \sum_{r=1}^{N} \frac{\left(\widehat{T}_{ro,t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o,t}^{k}\widehat{\tau}_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N}\left(\widehat{T}_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o',t}^{k}\widehat{\tau}_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N}\left(\widehat{T}_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o',t}^{k}\widehat{\tau}_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N}\left[\sum_{o'=1}^{N}\left(\widehat{T}_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o',t}^{k}\widehat{\tau}_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{108}$$

$$\widehat{P}_{d,t}^{k,ID} = \left[\sum_{r'=1}^{N}\left[\sum_{o'=1}^{N}\left(\widehat{T}_{r'o',t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o',t}^{k}\widehat{\tau}_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\right]^{-\frac{1}{\theta}} \tag{109}$$

$$\widehat{\varphi}_{od,t}^{k,ID} = \frac{\left(\widehat{T}_{ro,t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o,t}^{k}\widehat{\tau}_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N}\left(\widehat{T}_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o',t}^{k}\widehat{\tau}_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N}\left(\widehat{T}_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o',t}^{k}\widehat{\tau}_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N}\left[\sum_{o'=1}^{N}\left(\widehat{T}_{ro',t}^{k}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}_{o',t}^{k}\widehat{\tau}_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{110}$$

and the market clearing condition remains unchanged from equation 29. Changes in technology levels are given by:

$$\begin{aligned} T_{o,t'}^{k} - T_{o,t}^{k} &= \sum_{r} \delta_{ro,t}^{1-\rho}\gamma_{r,t}^{k}T_{r,t}^{k} \\ \implies \widehat{T}_{o,t'}^{k} &= 1 + \sum_{r} \delta_{ro,t}^{1-\rho}\gamma_{r,t}^{k}\frac{T_{r,t}^{k}}{T_{o,t}^{k}} \end{aligned} \tag{111}$$

and innovation levels remain unchanged from equation 36.

C.11 Proposition 4 (Regional and Aggregate Welfare)

Proof. I extend the welfare derivation and expression in Caliendo et al. (2019) to my setting, where: (i) preference shocks are correlated across markets; (ii) there is endogenous and microfounded innovation and technology diffusion, and (iii) the endogenous distribution of innovation and technology diffusion across markets drives parallel growth in all regions in the long run.

The expected worker value in equation (33) is given by:

$$V_{d,t}^{k,h} = U\left(C_{d,t}^{k,h}\right) + \frac{1}{\Upsilon}\log\left[\sum_{o}\sum_{s}\left(\sum_{n}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right] = \log\left(\frac{w_{d,t}^{k,h}}{P_{d,t}}\right) + \Phi_{d,t}^{k,h}.$$

I now express the option value $\Phi_{d,t}^{k,h}$ in terms of own-mobility shares:

$$\begin{aligned}
\Phi_{d,t}^{k,h} &= -\frac{1}{\Upsilon}\log\mu_{dd,t}^{kk} + \frac{\upsilon}{\Upsilon}\log\left[\sum_{n}\exp\left(\frac{1}{1+\zeta}V_{d,t'}^{k,n} - \kappa_{dd,t}^{kk,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right] \\
&= -\frac{1}{\Upsilon}\log\mu_{dd,t}^{kk} - \frac{\upsilon}{\Upsilon}\log\left(\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}\right) + \frac{\upsilon}{\Upsilon}\log\left[\exp\left(\frac{1}{1+\zeta}V_{d,t'}^{k,h}\right)^{\frac{\Upsilon}{\upsilon}}\right] \\
&= \frac{1}{1+\zeta}V_{d,t'}^{k,h} - \frac{1}{\Upsilon}\log\mu_{dd,t}^{kk} - \frac{\upsilon}{\Upsilon}\log\left(\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}\right).
\end{aligned}$$

Hence, expected worker value can be alternately expressed as:

$$V_{d,t}^{k,h} = \log\left(\frac{w_{d,t}^{k,h}}{P_{d,t}}\right) + \frac{1}{1+\zeta}V_{d,t'}^{k,h} - \frac{1}{\Upsilon}\log\mu_{dd,t}^{kk} - \frac{\upsilon}{\Upsilon}\log\left(\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}\right).$$

Iterating this equation forward, we have:

$$\begin{aligned}
V_{d,t}^{k,h} &= \sum_{t'\in\mathcal{T}_t^{\infty}}\left(\frac{1}{1+\zeta}\right)^{t'-t}\log\left(\frac{w_{d,t}^{k,h}}{P_{d,t}\left(\mu_{dd,t}^{kk}\right)^{1/\Upsilon}\left(\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}\right)^{\upsilon/\Upsilon}}\right) \\
&= \sum_{t'\in\mathcal{T}_t^{\infty}}\left(\frac{1}{1+\zeta}\right)^{t'-t}\left\{\underbrace{\log\left(\frac{w_{d,t}^{k,h}}{\widetilde{P}_{d,t}}\right)}_{\substack{\text{detrended}\\ \text{future value}}} - \underbrace{\log\left[\left(\mu_{dd,t}^{kk}\right)^{1/\Upsilon}\left(\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}\right)^{\upsilon/\Upsilon}\right]}_{\text{option value of movement}}\right\} + \underbrace{\frac{1+\zeta}{\zeta}\frac{1}{\theta}\sum_{k}\iota^{k}g^{k}}_{\text{long-run growth}}.
\end{aligned}$$

Denote $\acute{x}$ as the counterfactual of any variable x. My measure of the **welfare** impact in **market** (d,k,h) of an anticipated sequence of counterfactual changes in fundamentals from time $t=0$ is the compensating

variation in consumption for market (d,k,h) at $t=0$, $\log \delta_d^{k,h}$ given by:

$$\acute{V}_{d,0}^{k,h} = V_{d,0}^{k,h} + \sum_{t' \in \mathcal{T}_0^\infty} \left(\frac{1}{1+\zeta}\right)^{t'} \log \delta_d^{k,h}.$$

Rearranging this equation and substituting the expressions for expected worker value, we have:

$$\begin{aligned}
\log\left(\delta_d^{k,h}\right) &= \frac{\zeta}{1+\zeta}\left[\acute{V}_{d,0}^{k,h} - V_{d,0}^{k,h}\right] \\
&= \left(1-\frac{1}{1+\zeta}\right)\sum_{t' \in \mathcal{T}_0^\infty}\left(\frac{1}{1+\zeta}\right)^{t'} \log\left(\frac{\acute{w}_{d,t}^{k,h}}{w_{d,t}^{k,h}}\frac{\widetilde{P}_{d,t}}{\acute{\widetilde{P}}_{d,t}}\frac{\left(\mu_{dd,t}^{kk}\right)^{1/\Upsilon}\left(\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}\right)^{\upsilon/\Upsilon}}{\left(\mu_{dd,t}^{\acute{k}k}\right)^{1/\Upsilon}\left(\mu_{dd,t}^{k,h\acute{h}}|\mu_{dd,t}^{kk}\right)^{\upsilon/\Upsilon}}\right) + \frac{1}{\theta}\sum_k \iota^k\left(\acute{g}^k - g^k\right) \\
&= \sum_{t' \in \mathcal{T}_0^\infty}\left(\frac{1}{1+\zeta}\right)^{t'} \log\Bigg(\underbrace{\frac{\widehat{w_{d,t}^{k,h}}}{\widehat{\widetilde{P}_{d,t}}}}_{\substack{\text{change in}\\ \text{future}\\ \text{detrended}\\ \text{real wages}}} \underbrace{\frac{1}{\left(\widehat{\mu_{dd,t}^{kk}}\right)^{1/\Upsilon}\left(\widehat{\mu_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}}\right)^{\upsilon/\Upsilon}}}_{\text{change in option value of movement}}\Bigg) + \underbrace{\frac{1}{\theta}\sum_k \iota^k\left(\acute{g}^k - g^k\right)}_{\text{growth effects}},
\end{aligned}$$

where $\widehat{x}_{t'} = \frac{\dot{x}_{t'}}{\dot{x}_t}$ denotes the counterfactual change in any variable x.

Using this measure, I define **local and aggregate welfare** as population-weighted averages of welfare in the relevant markets:

$$\log(\delta_d) = \sum_{k,h}\frac{L_d^{k,h}}{\sum_{k,h} L_d^{k,h}}\log\left(\delta_d^{k,h}\right)$$

$$\log(\delta) = \sum_{d,k,h}\frac{L_d^{k,h}}{\sum_{d,k,h} L_d^{k,h}}\log\left(\delta_d^{k,h}\right).$$

□

D Model Extensions

My quantitative spatial growth model in the main text is deliberately parsimonious to capture the main drivers of the rising spatial concentration of innovation from the data. Nonetheless, the central feature of my spatial model is that I introduce endogenous innovation and integrate it with technology diffusion at the *idea* level, the fundamental unit of the Eaton-Kortum world. Thus, my model requires minimal assumptions and can flexibly incorporate other components in quantitative dynamic spatial models.

D.1 Dynamic Worker Sorting by Skill

To incorporate college-educated H and non-college educated S workers, equation 33-35 can be duplicated for each worker type. Equilibrium migration shares would now capture worker sorting patterns in the data as opposed to aggregate bilateral migration flows across both worker types. In the production of goods, labor

is now a composite of college-educated and non-college educated workers. For instance, one could assume a CES agggregate between both types of labor:

$$L_{o,t}^k = \left[\left(L_{o,t}^{H,k} \right)^{\frac{\varphi-1}{\varphi}} + \left(L_{o,t}^{S,k} \right)^{\frac{\varphi-1}{\varphi}} \right]^{\frac{\varphi}{\varphi-1}} \tag{112}$$

such that wages $w_{o,t}^k$ in the market clearing condition is now an aggregate of wages of each worker type, with no other required changes. Note that I drop the superscript G for production workers in this section of the appendix for notational clarity, since there is no discussion on innovation workers.

D.2 Multiple Factors of Production and Input-Output Loops

Input-output loops can be easily incorporated into the trade equilibrium at each t following Alvarez and Lucas (2007); Caliendo and Parro (2015). Instead of just using labor, production of each variety ν in each region is now given by a two-tier Cobb-Douglas constant returns to scale technology:

$$Y_{o,t}^k(\nu) = z_{o,t}^k(\nu) \left[(K_{o,t}^k)^{\psi} (L_{o,t}^k)^{1-\psi} \right]^{\chi} M_{o,t}^{1-\chi} \tag{113}$$

where $z_{o,t}(\nu)$ is the productivity drawn from the multivariate Fréchet distribution given by equation 13, $K_{o,t}$ is capital used in production, referring to commercial structures such as local buildings, $L_{o,t}$ is labor, and $M_{o,t}$ is intermediate inputs purchased from the final goods producer in the same region, ψ is the share of local structures in value added, and χ is the share of value added. The unit cost of an input bundle is given by:

$$x_{o,t}^k = \check{\Gamma} \left[(\check{r}_{o,t})^{\psi} \left(w_{o,t}^k \right)^{1-\psi} \right]^{\chi} {P_{o,t}^k}^{1-\chi} \tag{114}$$

where $\check{\Gamma}_o$ is a constant, $\check{r}_{o,t}$ is the rental rate of capital from local capitalists, and $P_{o,t}$ is also the price of the local industry aggregate of varieties. Replacing $w_{o,t}$ with $x_{o,t}$ in equation 18 yields the equilibrium trade shares.

Capital market clearing is given by:

$$\check{r}_{o,t} K_{o,t}^k = \frac{1-\psi}{\psi} w_{o,t}^k L_{o,t}^k \tag{115}$$

Since capital income is a constant multiple of production worker income, the combined capital and labor market clearing condition is still given by equation 29 in the main paper. With input-output loops of the form in equation 114 – where goods producers purchase the final good only in that sector[46] – intermediate good spending is a constant multiple of production worker and capital income:

$$X_{o,t}^I = \frac{1-\chi}{\chi} \left(w_{o,t}^k L_{o,t}^k + r_{o,t} K_{o,t}^k \right) = \frac{1-\chi}{\chi} \frac{1}{\psi} w_{o,t}^k L_{o,t}^k. \tag{116}$$

Thus, the combined capital and labor market clearing condition is still given by equation 29 in the main

[46] If goods producers purchase the final goods from all sectors, the market clearing condition will be slightly modified, as shown in Caliendo and Parro (2015), but remains highly tractable.

paper. All the other equations in the model also remain the same.

D.3 Capital Accumulation

Capital accumulation can be added following Kleinman et al. (2023). Apart from workers and local immobile firms, we can introduce local immobile capitalists. In each region, local immobile capitalists build durable local structures $K_{o,t}$ and rent to firms in different sectors at a nominal rate $\check{r}_{o,t}$. With their rental income, capitalists choose their consumption and investment to maximize intertemporal utility:

$$\check{V}_{o,t}^{K} = \mathbb{E}_t \sum_{s=0}^{\infty} \beta^{t+s} \frac{\left(C_{o,t+s}^{K}\right)^{1-1/\eta}}{1-1/\eta} \tag{117}$$

subject to their budget constraint:

$$r_{o,t} K_{o,t} = P_{o,t} \left(C_{o,t}^{K} + K_{o,t+1} - (1-\delta_{o,t}) K_{o,t} \right) \tag{118}$$

where rental income can be used for consumption (first term), or saving to increase future capital (last two terms). Per period consumption expenditures $P_{o,t} C_{o,t}^{K}$ are allocated via the same utility function as households.

The optimal consumption and saving decisions are given by:

$$C_{o,t}^{K} = \varsigma_{o,t} R_{o,t} K_{o,t} \tag{119}$$

$$K_{o,t+1} = (1-\varsigma_{o,t}) R_{o,t} K_{o,t} \tag{120}$$

$$\varsigma_{o,t}^{-1} = 1 + \beta^{\eta} \left(\mathbb{E}_t \left[R_{o,t+1}^{\frac{\eta-1}{\eta}} \varsigma_{o,t+1}^{-\frac{1}{\eta}} \right] \right)^{\eta} \tag{121}$$

where the consumption rate $\varsigma_{o,t}$ is defined recursively and the gross return on capital is $R_{o,t} \equiv 1-\delta+r_{o,t}/P_{o,t}$. The other equations in the model remain the same.

D.4 Amenities with Congestion and Agglomeration in Production

Amenities, explicit congestion forces, and agglomeration in production can be introduced following Allen and Arkolakis (2014). The instantaneous utility function is now given by:

$$U\left(C_{o,t}, B_{o,t}\right) = \log\left(B_{o,t} C_{o,t}\right) \tag{122}$$

with $B_{o,t} = \check{B}_{o,t} L_{o,t}^{-\xi}$ where $\check{B}_{o,t}$ are fundamental amenities in region o at time t and ξ captures congestion forces. The goods production function in equation 113 now becomes:

$$Y_{o,t}^{k}(\nu) = z_{o,t}^{k}(\nu) (L_{o,t}^{k})^{\check{\alpha}} \left[(K_{o,t}^{k})^{\psi} (L_{o,t}^{k,G})^{1-\psi} \right]^{\chi} M_{o,t}^{1-\chi} \tag{123}$$

where $\check{\alpha}$ captures agglomeration in production. All the other equations in the model remain the same.