The Geography of Innovation in the United States

Weiliang Tan*
Cornell University

November 24, 2024

JOB MARKET PAPER
Click here for the latest version

Abstract

One of the most striking trends in the study of innovation is its rising spatial concentration in the United States, exemplified by the emergence of high-tech clusters like Silicon Valley in recent decades. Why is innovation increasingly concentrated geographically, and beyond overall innovation levels, what are the broader implications of this concentration on spatial and aggregate growth? Using comprehensive data on patents, firms, and inventors from 1976-2018, I find that innovation became more concentrated in high-skill cities only from 1990 onward, 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 the ICT service sector and concentrated in high-skill cities. Second, firms initially concentrated in high-skill cities also produced more non-ICT patents due to spillovers from ICT innovation and reductions in communication costs enabled by ICT, which allowed these firms to expand production to lower-cost regions and enhanced the profitability of new ideas. To better understand the mechanics of innovation across space and its consequences for macroeconomic growth, I develop a novel model of spatial growth with endogenous and directed innovation and technology diffusion. 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; 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 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 question in spatial economics is what drives the distribution of economic activity across space and what the local and aggregate consequences of this distribution are. This paper addresses the question in the context of innovation – the creation of ideas – a fundamental economic activity in modern societies, especially in the developed world.

One of the most striking facts in the study of innovation is its rising spatial concentration in the United States, exemplified by the rise of high-tech clusters like Silicon Valley in recent decades. 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 and uses estimated agglomeration spillovers to demonstrate that this spatial concentration leads to greater aggregate innovation than if inventors were evenly distributed across space. More fundamentally, however, little is known about how these high-tech clusters originated and why their emergence 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 and core mechanisms driving the spatial concentration of innovation is challenging due to the wide range of potential explanations, but crucial for developing effective placed-based policies that 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 high-tech clusters emerged. 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 significantly 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 rise 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 annual 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 the ICT service sector – comprising Software Publishers; Telecommunications; 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 annual share of ICT patents, which are more concentrated in high-skill cities, accounts for 53% of the overall increase in the concentration of innovation in high-skill cities from 1990 to 2018. Additionally, the ICT service sector – 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 10% increase in the firm's initial elasticity of patents per capita with respect to the 1990 college ratio between 1988 and 1990 is associated with a 2ppt greater increase in non-ICT patents from 1990 to 2000. 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 increased the profitability of new ideas – a phenomenon I term the *asymmetric scale effect*. Specifically, after controlling for a firm's ICT patents in the respective year, 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, the 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 10% increase in the firm's elasticity of employment per capita with respect to the 1990 college ratio between 1988 and 1990 is associated with a 0.4% greater increase in the number of CZs with the firm's establishments from 1990 to 2000. 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 over the same period. These elasticities indicate that firms initially concentrated in high-skill cities were more likely to expand to lower-cost regions after 1990 but not before, providing first evidence of the asymmetric scale effect: a novel productivity advantage of cities following the ICT shock.

My empirical findings suggest that the ICT shock from 1990 onwards 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 the ICT service sector 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 (the asymmetric scale effect). To formalize and better understand each of these mechanisms, I develop a quantitatively-oriented theory of spatial growth that centers on endogenous, directed, and microfounded innovation, integrated with technology diffusion and worker mobility. My theory builds on modeling techniques from Eaton and Kortum (2001) and Lind and Ramondo (2024), integrating endogenous innovation and technology diffusion at the level of individual ideas, the fundamental unit in the Eaton-Kortum structure. Consequently, my model characterizes the degree of *colocation* between innovation and production. It also explains why high-wage regions disproportionately benefit from a uniform rise in bilateral diffusion speeds nationwide – representing reduced communication costs from the ICT shock – due to their greater gains in idea market access, what I call the *asymmetric scale effect*. These equilibrium results cannot be obtained in existing spatial growth models, where innovation is either highly stylized or modeled as independent of technology diffusion.

My model features two sectors, ICT and non-ICT. In each region and sector, inventors receive ideas from a Poisson process, with arrival rates determined by: (i) the regional college ratio; (ii) sector-specific, time-varying research productivity (capturing the compositional shift of innovation towards ICT); (iii) agglomeration economies in innovation (capturing local spillovers from ICT to non-ICT innovation), and; (iv) region-specific, time-varying research productivity (capturing factors unexplained by the model). Each idea corresponds to the production of a specific good from the unit interval. Once an idea is discovered in one region, its diffusion to other regions is governed by independent Poisson processes, with arrival rates proportional to the intensity of within-firm connections between the corresponding regions, thereby capturing firm spatial expansion. Each idea has two additional components – modeled as marks on each Poisson process – 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 levels in all regions and diffusion speeds in all region-pairs. Additionally, goods productivity draws are correlated across regions due to common idea discovery locations, generating more realistic trade patterns. Notably, the equilibrium trilateral trade shares are determined by the product of idea diffusion and idea market shares, reflecting the spatial mechanics of innovation documented in my empirical findings. The idea diffusion shares characterize the degree of colocation between innovation and production, while the idea market shares show that a uniform increase in diffusion speeds across all region-pairs disproportionately benefits high-skill cities due to a greater increase in access to lower-cost regions for producing goods derived from new ideas.

While the aggregate idea diffusion and idea market shares capture the spatial mechanics of innovation, a model of endogenous innovation requires that agents have incentives to innovate. In my model, there is a unit continuum of firms in each market (defined by region and sector), which hire inventors and own 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. Over time, workers make dynamic decisions about moving across regions and sectors, as well as between production and research, based on wages from innovation and production. Along the transition path, the ratio of real wages from innovation across regions characterizes the incentives for workers mobility and, consequently, the spatial direction of innovation.

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.

This paper makes two central contributions across different fields in economics. 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 emergence of these high-tech clusters remains unanswered. I identify the rise of high-tech clusters from 1990 onward as having primarily occurred in high-skill cities, with the ICT shock playing two distinct roles: first, by driving a shift in innovation toward ICT, which is colocated with the ICT service sector and already concentrated in these regions; and second, by enabling firms initially concentrated in high-skill cities to produce more non-ICT innovation through spillovers from ICT and the asymmetric scale effect of reduced communication costs.

The third mechanism in my explanation 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. The first mechanism in my explanation – the colocation of ICT innovation and the ICT service sector – 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 distinguishing between R&D plants with plants in the ICT service sector, classified under their list of professional service industries. Additionally, I examine the intensive margin of colocation through employment shares and investigate how colocation influences the aggregate spatial concentration of innovation. More broadly, my explanation of how sectoral shifts in innovation shape its aggregate geography connects to the macro-development literature on structural change, which typically abstracts from both innovation and spatial considerations (see for e.g. Herrendorf et al., 2014). Most closely related is Eckert and Peters (2023), which examines how the geography of development interacts with structural change in production but abstracts from innovation.

Second, I develop a novel model of spatial growth that integrates endogenous and directed innovation with technology diffusion and worker mobility, thereby contributing to the quantitative trade, spatial, and macroeconomic literatures. Building on modeling techniques from Eaton and Kortum (2001) and Lind and Ramondo (2023a, 2024), I model innovation and technology diffusion at the level of individual ideas, ensuring maximum tractability. Unlike these models, however, innovation in my framework is fully endogenous and depends not only on equilibrium trade but also the entire technology diffusion network. Consequently, I provide the first analytical characterization of the degree of colocation between innovation and production, as well as the first illustration of the asymmetric scale effect – how a uniform rise in bilateral diffusion speeds disproportionately benefits high-wage regions. 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) incorporate endogenous innovation but exclude 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 diffusion 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.

Additionally, unlike the quantitative trade and innovation literature, my model incorporates dynamic worker mobility with frictions and provides an analytical characterization of both the balanced growth path – the primary focus of this literature – and the transition path. By characterizing the transition path in response to shocks in economic fundamentals, my work aligns with the class of quantitative dynamic spatial models with worker migration (Caliendo et al., 2019), capital accumulation (Kleinman et al., 2023), and knowledge diffusion (Cai et al., 2022). I extend this class of models by introducing endogenous and directed innovation while maintaining tractability, allowing for the potential integration of these other mechanisms. Desmet et al. (2018) is 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 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. Furthermore, my analytical characterization of the spatial direction of innovation on the transition path extends 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 US: 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 emergence of Silicon Valley-like clusters in recent decades. The fundamental drivers of this trend has, however, remained elusive in over two decades of research on the geography of innovation. Understanding why innovation became increasingly concentrated in space requires a careful examination 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 innovation in the United States; (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 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); Dreisigmeyer et al. (2018). Within firms, I alternately assign inventors to the CZ of their home city or to 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 Database. See Appendix A for more details.

2.1 Fact 1: Innovation became more spatially concentrated in high-skill cities from 1990 onwards but not before

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 shares of total patents and population in the US. I then rank CZs by their patent-to-population share ratio and plot the cumulative sum of patent shares against 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].

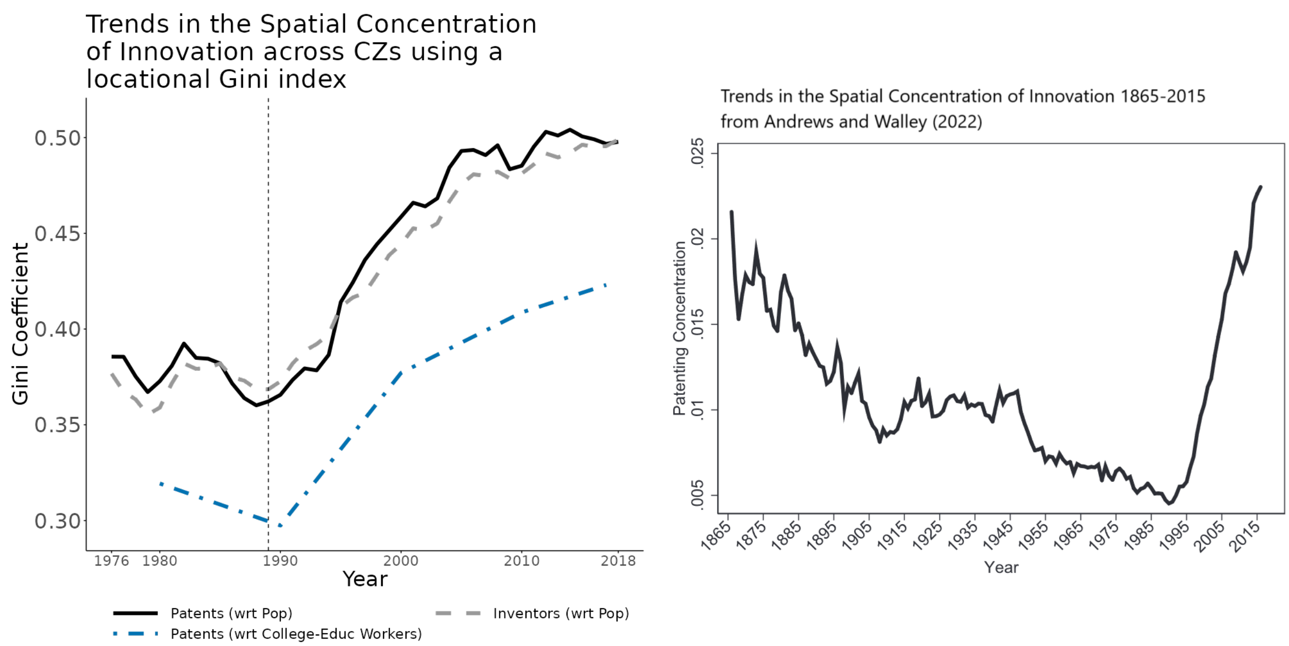


Figure 1: Trends in the spatial concentration of innovation across CZs. The left graph plots trends in the spatial concentration of innovation from 1976-2018, measured with respect to population and college educated workers using a locational Gini index (Krugman, 1991) and USPV data. For each CZ, I calculate 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 graph shows trends in the spatial concentrated of innovation from 1865-2015 documented by Andrews and Whalley (2021).

[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:

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

Using this index, the solid black line in the left graph 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 reflecting agglomeration economies in innovation, as documented by Moretti (2021). 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 from 1990. This finding suggests that the rising spatial concentration of innovation cannot be attributed to the geographical sorting of workers by skill from 1980-2000, as documented in Moretti (2013); Diamond (2016). In Figure 13 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 graph presents trends in the spatial concentration of innovation from 1865 to 2015, as documented by Andrews and Whalley (2021) using an alternative patent dataset[2]. This figure also reveals a sharp rise in the spatial concentration of innovation beginning in 1990.

The locational Gini curve also offers a natural, annual measure of patenting activity for each CZ that allows for meaningful comparisons over time: the patent share per unit population share. 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. Changes in patent share per unit population share 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. To the best of my knowledge, this is the **first local measure of changes in the geography of innovation over time**[3]. Figure 2 depicts the geography of changes in patent share per unit population share 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.

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

[3]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.

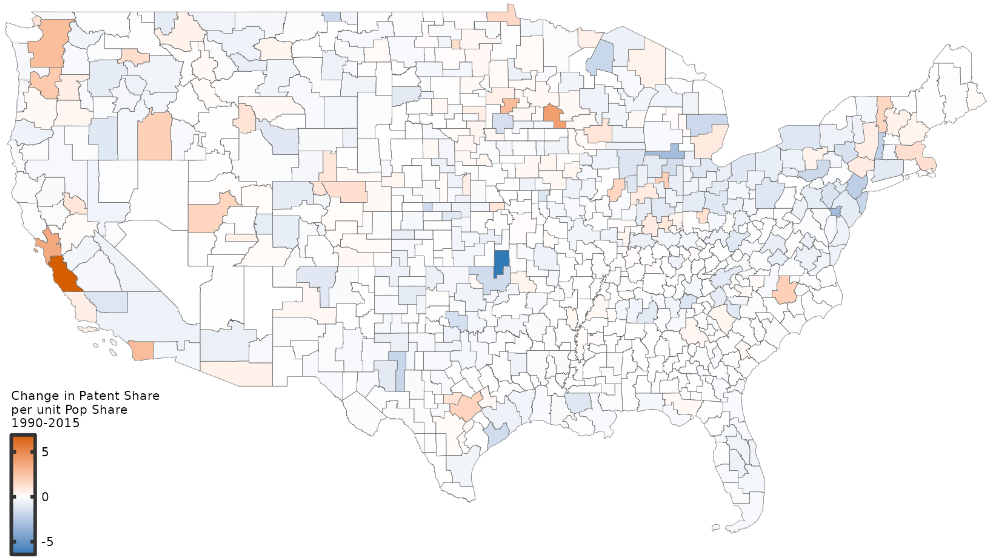


Figure 2: Changes in the geography of innovation intensity from 1990-2015, measured using changes in patent share per unit population share on a pseudo-log scale.

Using this measure, Figure 3 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 patent share per unit population share 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 patent share per unit population share 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[4] with respect to the 1990 college ratio using the following PPML regression:

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

where r represents regions and t denotes years. The estimated annual aggregate elasticities α_t serve as the central focus for the remainder of my empirical analysis. The right graph of Figure 3 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 14 in Appendix B.1, I provide first evidence that worker sorting by skill in the US predominantly occurred in the 1980s and did not continue after 1990. Additionally, Figure 15 in Appendix B.1 plots trends in the PPML 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.

[4] This is equivalent to patent share per unit population share with year fixed effects.

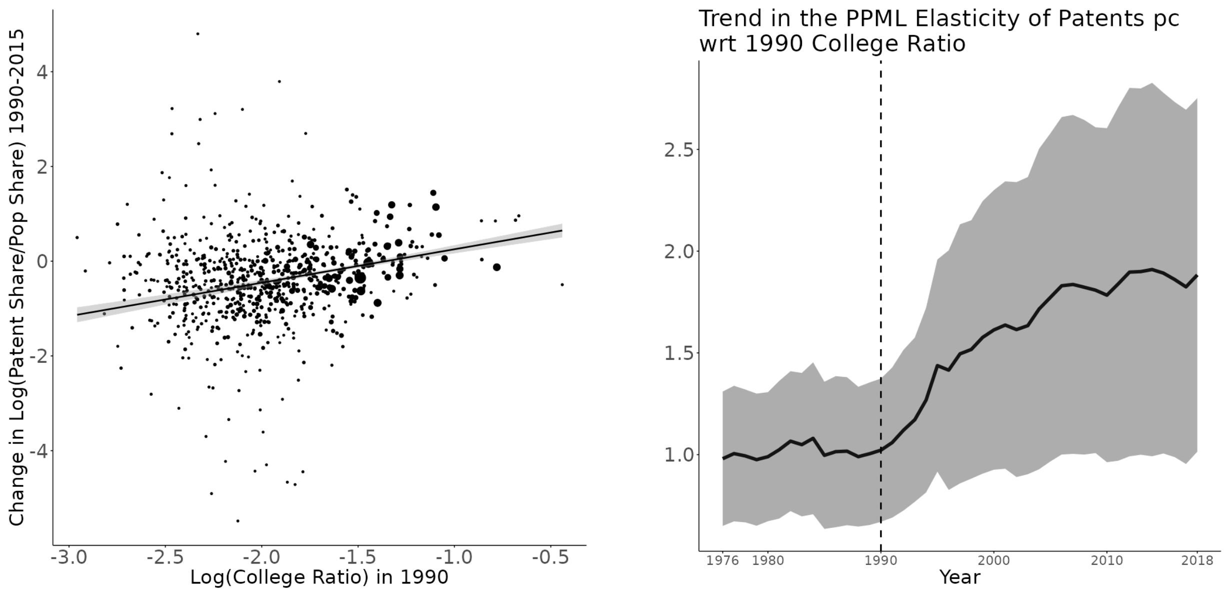


Figure 3: The geographical sorting of patents per capita into high-skill cities from 1990. The left graph shows how percentage changes in patent share per population share from 1990-2018 relate to the initial college ratio across CZs. Specifically, it plots changes in log patent share per population share between five-year averages around 1990 and 2015 against the five-year average of log college ratio around 1990, with each CZ weighted by its population. The right graph plots trends in the annual PPML elasticity of CZ patents per capita with respect to the 1990 college ratio. The confidence intervals in both graphs reflect the statistical uncertainty in the slope estimates using heteroskedastic-robust standard errors.

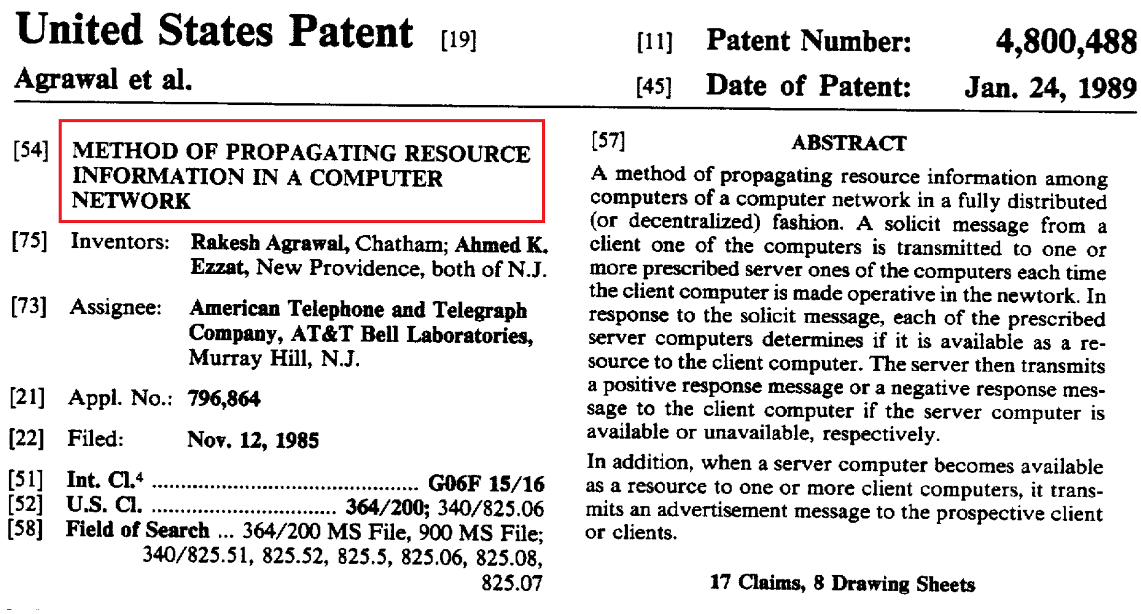
2.2 The Rapid Rise of Information and Communication Technologies (ICT shock)

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 significant compositional shift in economic activity towards the ICT sector. Kelly et al. (2021) identify eight breakthrough patents[5] 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 4 hightlights the content of the first of these breakthrough patents, which introduced a method for propagating resource information in a computer network.

More concretely, I define this direct aspect of the ICT shock as a sharp rise in the annual share of innovation and production in the ICT sector from around 1990. The left graph of Figure 5 plots trends in the annual number of patents for each technology field[6]. While the number of patents has generally increased over time across all fields, the growth in ICT patents from 1990 to 2018 is particularly pronounced. This is reflected in the sharp rise in the annual share of patents in ICT from approximately 8.5% in 1990 to 33% in 2018,

[5]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

[6]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 4: Contents of the first of eight breakthrough patents in ICT identified by Kelly et al. (2021)

as shown in the middle graph. Shifting from innovation to production, the right graph plots trends in the national employment share of the ICT service sector using harmonized County Business Patterns data from Eckert et al. (2020). 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). The graph shows a sharp increase in the employment share in the ICT service sector from about 1.4% in 1990 to 2.0% in the early 2000s.

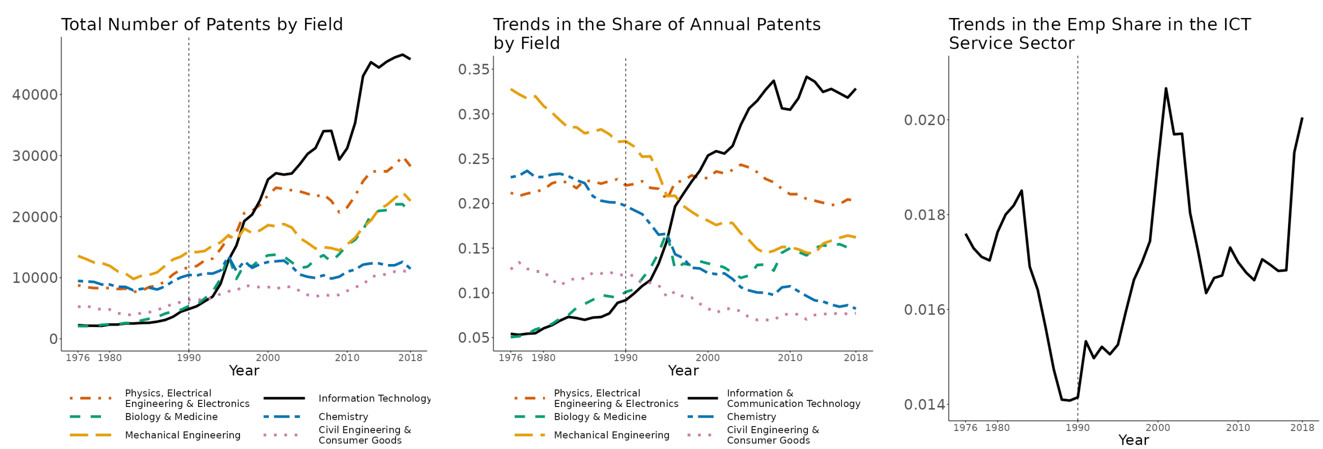


Figure 5: Trends in the number of patents (left) and share of annual patents (middle) by technology field, as well as the national employment share in the ICT service sector (right).

Second, the ICT shock significantly reduced communication costs across regions. Greenstein (2015); 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.

2.3 Fact 2 (Direct Effect of ICT): ICT innovation is more concentrated in high-skill cities relative to other fields because it is colocated with the ICT service sector

The compositional shift of innovation toward ICT increased the aggregate spatial concentration of innovation, as ICT patents are more concentrated in high-skill cities compared to patents in other fields. Figure 6 illustrates this disparity: both the Gini coefficient of patents per capita (left graph) and the PPML 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, even before 1990.

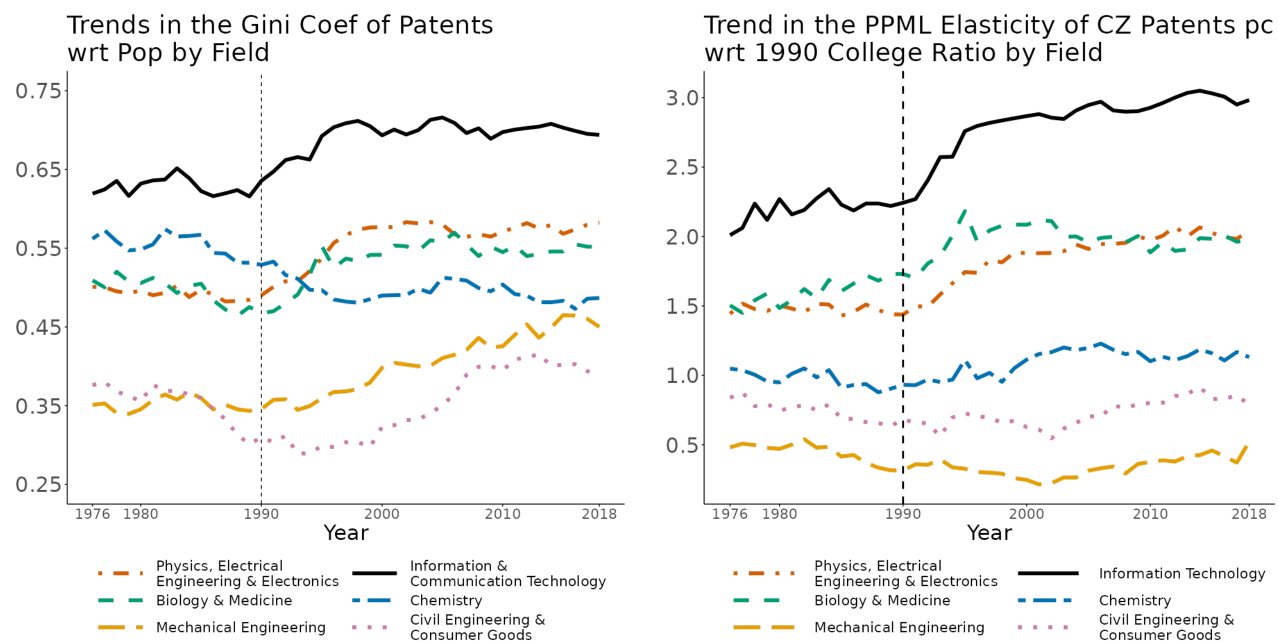


Figure 6: Trends in the Gini coefficient (left) and PPML elasticity with respect to 1990 college ratio (right) of patents per capita by technology field.

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

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

where $\alpha_{k,t}$ is the annual PPML 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 varialbe 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 PPML elasticity of patents per capita against the 1990 college ratio from a weighted mean of the field-specific PPML elasticities. The left graph of Figure 7 plots trends in this decomposition, while the right graph further decomposes the cross-field component into contributions from individual fields. These graphs show that 53% of the increase in the overall elasticity is explained by compositional shifts in patents towards ICT[7]. In Appendix B.3.1, I decompose the rise in the overall Gini coefficient of patents per capita from 1990-2018 and find that 52% of the increase is similarly driven by the rising annual share of ICT patents.

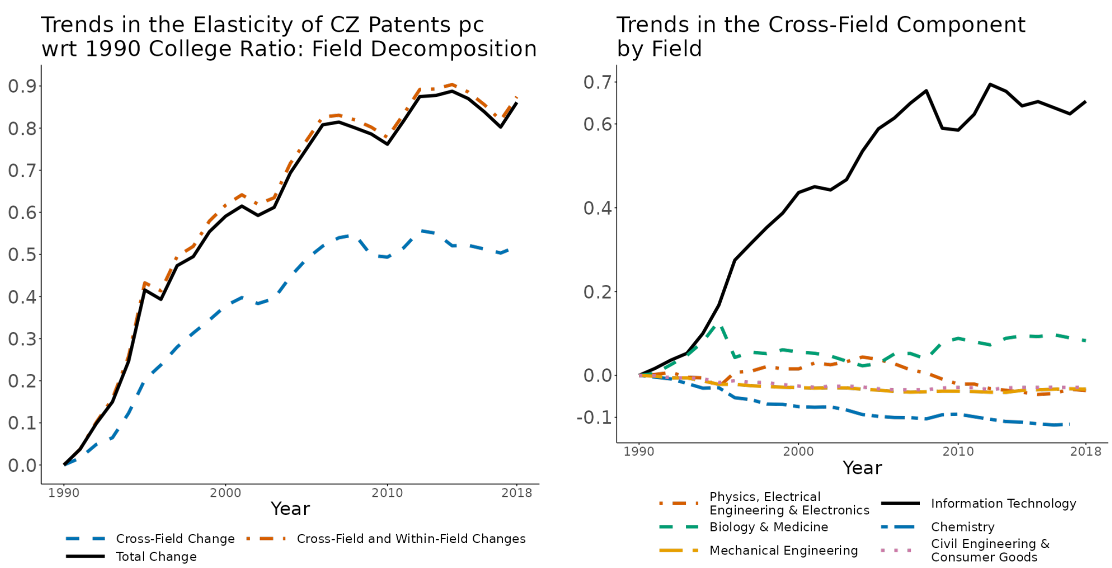


Figure 7: Trends in the decomposition of the aggregate PPML elasticity of CZ patents per capita with respect to the 1990 college ratio. The left graph decomposes the elasticity into within versus cross field components, while the right graph further breaks down the cross-field component into contributions from individual fields.

[7] More precisely, 60% of the increase in the overall elasticity is attributed to cross-field changes, of which 91% arises from the rising share of ICT patents and 9% from the rising share of Biology patents.

An underlying mechanism that explains the greater concentration of ICT innovation in high-skill cities relative to other fields is its **colocation with the ICT service sector**. The left table of Figure 8 provides evidence of this mechanism from the following regression:

$$\text{ICT Patents pc}_{r,t} = \beta \cdot \text{ICT Employment Share}_{r,t} \times 1(\text{Year} \geq 1990) + \text{Year FE} + \text{CZ FE} + \varepsilon_{r,t}. \quad (2)$$

where r represents CZs and t represents years. Specifications (1) and (3), which include year fixed effects, show that before 1990, a 10% increase in the employment share of the ICT service sector 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 highlight the stronger colocation of the ICT service sector with ICT innovation compared to non-ICT innovation. In the post-1990 subsample, both estimates increased, and the widening gap between them reflects an even stronger colocation between the ICT service sector and ICT innovation after 1990. Specifications (2) and (4), which include both year and CZ fixed effects, 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. Additionally, I run this regression at the firm-CZ-year level using US Census LBD data with a saturation of fixed effects, but have yet to disclose the estimates. These results, summarized in the left table of Figure 8, extend the findings of Fort et al. (2020) by offering an ICT sector-specific analysis of the colocation of innovation and production. Together, these findings underscore that colocation is a critical mechanism driving the greater concentration of ICT innovation in high-skill cities relative to non-ICT innovation.

Dependent Variable:	ICT patents pc		Non-ICT patents pc	
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***
× From 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*

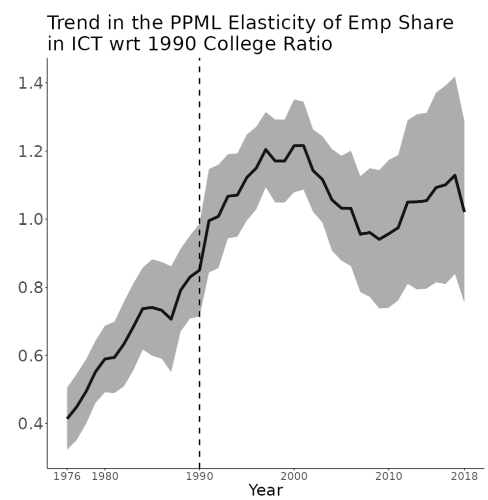


Figure 8: Colocation of ICT patents with employment share of the ICT service sector, which is concentrated in high-skill cities. The left table shows the correlation between ICT/non-ICT patents per capita and the employment share of the ICT service sector with year and CZ fixed effects. Each observation is at the CZ-year level, weighted by CZ population. The right graph plots trends in the annual PPML elasticity of the CZ employment share of the ICT service sector with respect to the 1990 college ratio.

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 graph plots trends in the annual PPML elasticity of the CZ employment share in the ICT service sector with respect to the 1990 college ratio, captured by the estimated γ_t coefficients from the following regression:

$$\text{ICT Emp Share}_{r,t} = \exp\left(\gamma_t \cdot \text{Log 1990 College Ratio}_r \times \text{Year}_t + \text{Year FE} + \epsilon_{r,t}\right). \tag{3}$$

where r represents commuting zones (CZs) and t represents years. These elasticities are positive even before 1990, with a 10% increase in the 1990 college ratio across CZs associated with a 4-8% higher employment share in the ICT service industry in various years prior to 1990. This finding indicates that the ICT service sector was already concentrated in high-skill cities before the ICT shock, laying the foundation for the subsequent rise in ICT innovation in these regions. Moreover, the annual elasticity of the ICT employment share with respect to the 1990 college ratio increased steadily over time, peaking around 2000. This trend highlights the growing concentration of the ICT service sector in high-skill cities during this period. Given the colocation of ICT innovation and production, this rising concentration of the ICT service sector contributed to the increasing concentration of ICT innovation in these cities, accounting for an additional 13% of the overall rise in the patent concentration in high-skill cities between 1990 and 2000.

2.4 Fact 3 (Indirect Effect of ICT): Firms initially concentrated in high-skill cities drove non-ICT patent growth after 1990 through ICT spillovers and expansion to lower-cost regions, enhancing the profitability of new ideas

An important question that remains 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 from 1990 onwards. To investigate this, I first document trends in the spatial concentration of innovation within each field. Figure 9 replicates the trends in the Gini coefficient and PPML elasticity by field from Figure 6, normalizing the 1990 levels to zero for clearer comparison over time. The left graph shows that the locational Gini coefficient increased for all fields except Chemistry between 1990 and 2018. Notably, the spatial concentration of patents in ICT, Physics, Electrical Engineering & Electronics, and Biology & Medicine rose sharply during the 1990–2000 period. These fields predominantly concentrated in high-skill cities, as evidenced by the right graph: the annual PPML elasticity of CZ patents per 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[8]. 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

[8] 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.

during this period, though to a smaller extent compared to the sharper rises observed for ICT, Physics, Electrical Engineering & Electronics, and Biology & Medicine.

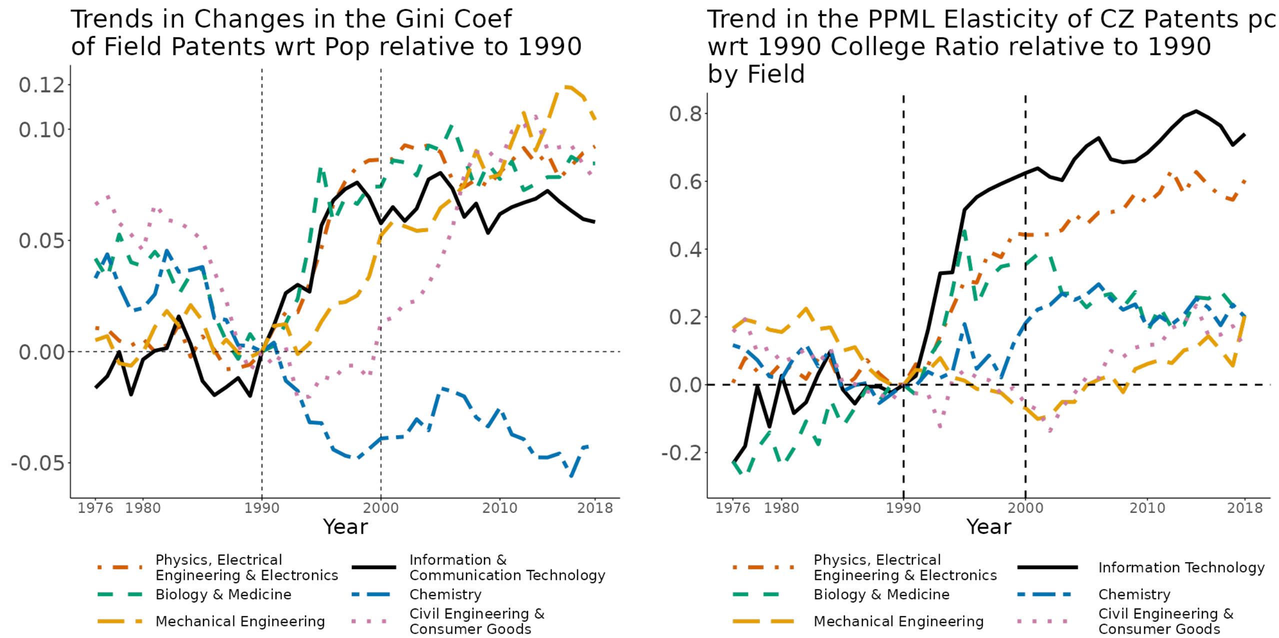


Figure 9: Trends in changes relative to 1990 of the locational Gini coefficient (left) and elasticity with respect to 1990 college ratio (right) of patents per capita by technology field.

The rising concentration of non-ICT innovation in high-skill cities since 1990 can be attributed to various factors, including the entry of new firms, the geographic expansion of existing firms, or compositional changes in patent shares across firms. Identifying the primary drivers is challenging, given the absence of an exhaustive list of contributing factors. To make progress, I first decompose the rising spatial concentration of non-ICT patents in high-skill cities into within-firm and across-firm components. This approach helps disentangle whether the observed trends are primarily driven by changes within patenting activity within firms, or shifts in the composition of patenting firms across regions. The decomposition is formalized as:

$$\alpha_{t^*} - \alpha_{1990} = \sum_{t=1991}^{t=t^*} \Delta\alpha_t = \sum_{t=1991}^{t=t^*} \left[\underbrace{\sum_f \overline{\alpha}_{f,t} \Delta s_{f,t}}_{\substack{\text{changes in firm} \\ \text{composition}}} + \underbrace{\sum_f \overline{s}_{f,t} \Delta \alpha_{f,t}}_{\substack{\text{within-firm} \\ \text{changes}}} + \underbrace{\Delta\left(\alpha_t - \sum_f s_{f,t}\alpha_{f,t} \right)}_{\substack{\text{changes in the} \\ \text{colocation of firms}}} \right], \tag{4}$$

where $\alpha_{f,t}$ is the annual PPML elasticity of CZ patents per capita in firm f 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. Analogous to the field decomposition, the first term captures the impact of changes in the composition of patenting firms, the second term reflects changes in the spatial concentration of patents within firms in high-skill cities, and the third term accounts for the residual. The left graph of Figure 10 plots trends from this decomposition, showing that the rising concentration of non-ICT innovation in high-skill cities is entirely driven by compositional changes across firms.

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 graph of Figure 10 presents trends from this decomposition, revealing that most of the increase is driven by firms that began patenting before 1990[9].

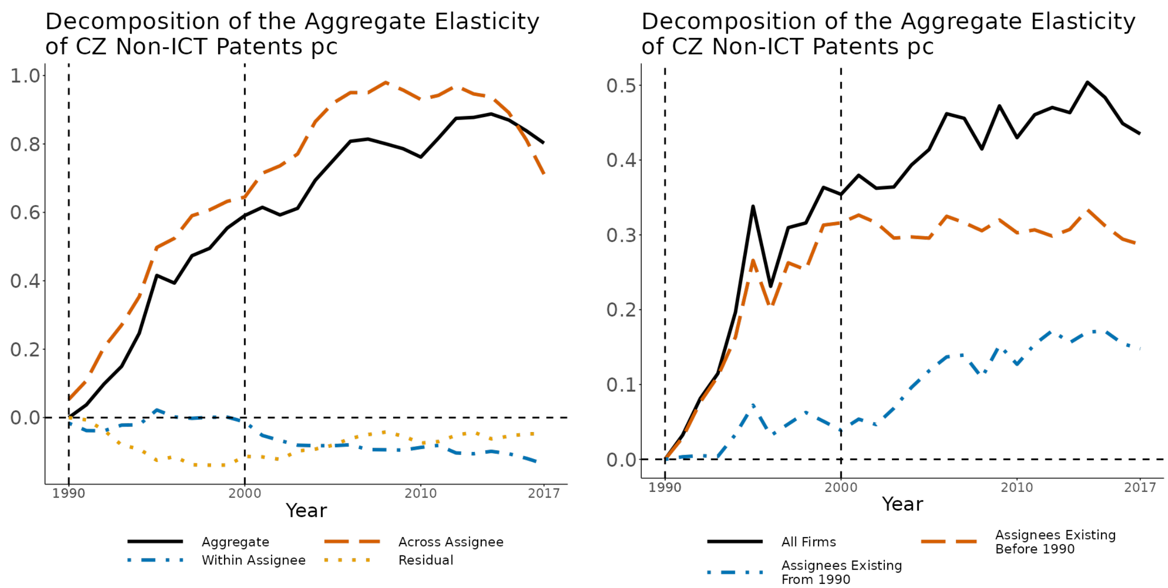


Figure 10: Decomposition of the aggregate PPML elasticity of CZ non-ICT patents per capita with respect to the 1990 college ratio into within versus across assignee (left) and between firms existing before 1990 versus new firms (right).

More directly, I find that firms initially concentrated in high-skill cities from 1988 to 1990 produced significantly more patents after 1990, but not before. Specifically, I estimate the δ_t coefficients from the following PPML regression:

$$\begin{aligned}\text{Non-ICT Patents}_{f,t} = \exp\big(&\delta_t \cdot \text{1990 Patent Elasticity}_f + \text{ Year FE } + \text{ Firm FE} \\ &+ \text{1990 Patent Elasticity}_f + \text{ 1990 Patents}_f + \text{1990 Patents CZ No}_f + \epsilon_{f,t}\big)\end{aligned} \tag{5}$$

where 1990 Patent Elasticity$_f$ is firm f's average annual PPML elasticity of patents per capita with respect to the 1990 college ratio from 1988 to 1990, 1990 Patents$_f$ is the firm's average number of annual patents from 1988 to 1990 (a proxy for firm size), and 1990 Patents CZ No$_f$ is the firm's average annual number of CZs of their inventors across all patents from 1988 to 1990 (a proxy for firm spatial scope). Figure 11 plots the δ_t coefficients, showing that between 1990 and 2000, a one unit increase in the firm's 1990 patent elasticity with respect to the college ratio is associated with 20 percentage point greater increase in the number of non-ICT patents.

[9]Note that I attribute both the rising share of patents from firms that began patenting from 1990 and the rising concentration of patents within this group to the contribution of assignee entry (i.e. assignees existing from 1990).

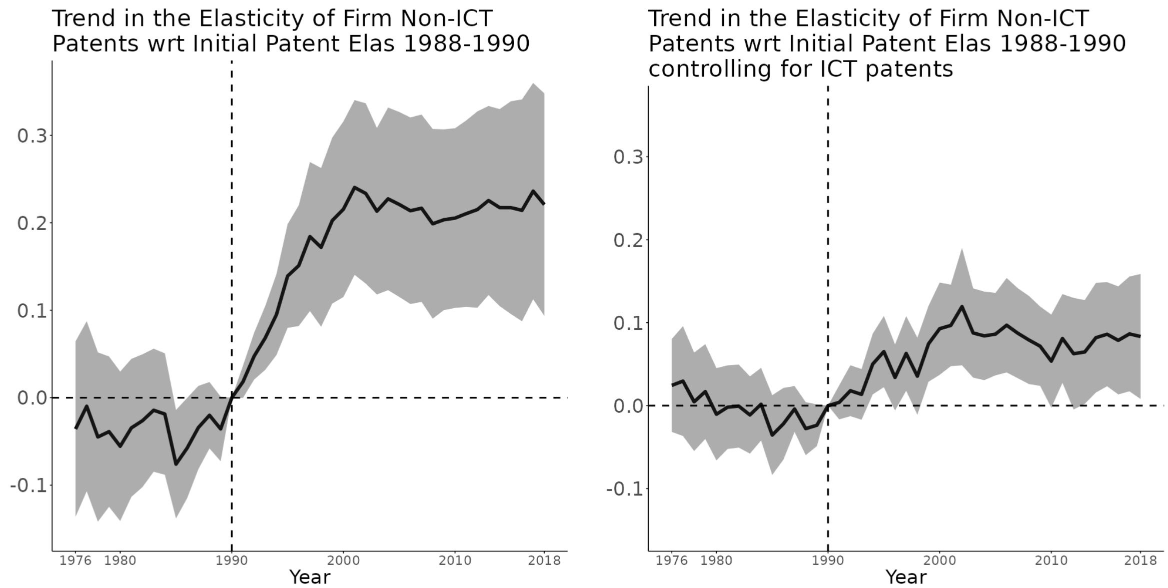


Figure 11: Trends in the semi-elasticity of non-ICT patents with respect to a firm's initial concentration of patents in high-skill cities (1988-1990) (left), and after controlling for the firm's ICT patents in the respective year (right).

Why did firms initially concentrated in high-skill cities produce more non-ICT patents after 1990? One mechanism is **spillovers from ICT to non-ICT innovation**. The right graph of Figure 11 plots the δ_t coefficients from equation 5, now incorporating a time-varying control for the firm's ICT patents in the respective year. The results show that the increase in firm-level non-ICT patents 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. These spillovers were predominantly concentrated in high-skill cities, where the ICT boom occurred.

Another key mechanism is that **firms initially concentrated in high-skill cities disproportionately expanded to new, lower-skill production locations after 1990**. The left graph of Figure 12 plots the trend in the PPML semi-elasticity κ_t of the number of CZs with the firm's establishments with respect to the firm's initial employment elasticity. These semi-elasticities are estimated from the following specification:

$$\begin{aligned}\text{No of CZs of Plants}_{f,t} = \exp\big(&\kappa_t \cdot 1990 \text{ Emp Elasticity}_f + \text{ Year FE } + \text{ Firm FE} \\ &+ 1990 \text{ Emp Elasticity}_f + 1990 \text{ Emp}_f + 1990 \text{ No of CZs of Plants}_f + \epsilon_{f,t}\big)\end{aligned} \tag{6}$$

where 1990 Emp Elasticity$_f$ represents firm f's average annual PPML elasticity of employment per capita with respect to the 1990 college ratio, 1990 Emp$_f$ denotes the firm's average annual employment, and 1990 No of CZs of Plants$_f$ captures the firm's average annual number of CZs containing its plants. These annual averages are calculated over the period from 1988 to 1990. The trend 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 4ppt greater increase in the number of CZs containing the firm's plants in 2000. This effect is not an artifact of the regression specification: the right graph plots trends when the dependent variable in equation 6 is replaced with the firm's total employment, and shows no significant increases after 1990.

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. Specifically, I estimate the υ_t coefficients from the following specification:

$$\begin{aligned}\text{Emp Elasticity}_{f,t} = \upsilon_t \cdot \text{1990 Emp Elasticity}_f + \text{Year FE } + \text{Firm FE}\\ + \text{1990 Emp}_f + \text{1990 No of CZs of Plants}_f + \epsilon_{f,t}\end{aligned} \tag{7}$$

where Emp Elasticity$_{f,t}$ is firm f's annual PPML elasticity of employment per capita with respect to the 1990 college ratio and the explanatory variables are identical to the ones in equation 6. The middle graph of Figure 12 plots 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. 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. 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.

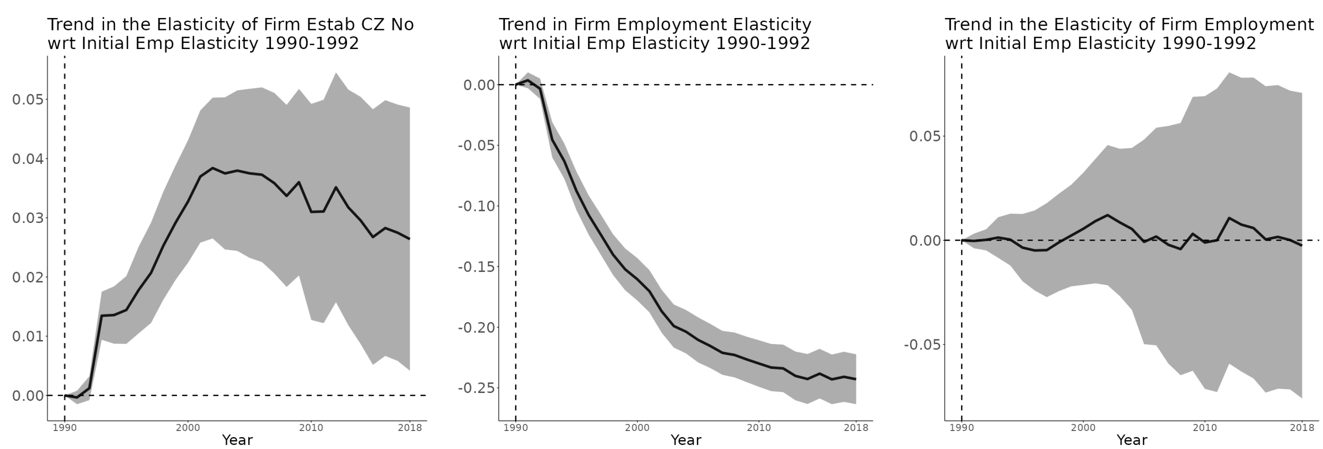


Figure 12: Trends in the PPML semi-elasticity of the number of CZs containing the firm's establishments (left) and the firm's total employment (right) with respect to the firm's concentration of employment in high-skill cities from 1988 to 1990, along with trends in the correlation between the firm's annual employment elasticity and its initial concentration of employment in high-skill cities (middle).

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

My empirical findings suggest that the ICT shock around 1990 accounts for most of the rising spatial concentration of innovation in high-skill cities after 1990 through three distinct mechanisms. The first is a direct effect: a compositional shift of innovation towards ICT, which is ***colocated*** with the ICT service sector and more concentrated in high-skill cities relative to innovation in other fields. In addition, firms initially concentrated in high-skill cities produced more non-ICT patents from 1990, likely due to two indirect effects of the ICT shock: (i) ***spillovers*** from ICT to non-ICT innovation, and (ii) geographic expansion into lower-cost regions, facilitated by reduced communication costs, which increased the profitability of new ideas – a phenomenon I term the ***asymmetric scale effect***.

To formalize and better understand these mechanisms, I develop a model of spatial growth with endogenous and directed innovation, technology diffusion, and worker mobility. My theory builds on modeling techniques from Eaton and Kortum (2001) and Lind and Ramondo (2024), integrating endogenous innovation and technology diffusion at the level of individual ideas. Consequently, my framework provides the first analytical characterization of the degree of colocation of innovation and production, as well as the first formal explanation for the asymmetric scale effect: high-skill cities benefit disproportionately from a uniform increase in bilateral diffusion speeds nationwide (capturing the reduction in communication costs from the ICT shock) due to greater gains in idea market access. These mechanisms align closely with my empirical findings and cannot be derived from existing spatial growth models, where innovation is either highly stylized or modeled as independent of technology diffusion.

In my model, the productivity distribution of goods – and consequently the trade shares – is 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. The model features two sectors: ICT and non-ICT. In each region and sector, innovation workers generate new ideas through a Poisson process, where the arrival rate is influenced by several key factors. These include sectoral productivity in innovation, which captures compositional shifts in innovation toward ICT (the direct effect of the ICT shock), and agglomeration economies, including spillovers from ICT to non-ICT innovation (the first indirect effect of the ICT shock). Once an idea is discovered in a particular region, its diffusion to other regions is governed by independent Poisson processes. I alternately assume that the diffusion rate is either homogeneous across all region-pairs, or proportional to the intensity of within-firm connections between the corresponding regions. Both formulations capture the asymmetric scale effect from falling communication costs (the second indirect effect of the ICT shock). With additional assumptions, 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. This scale parameter evolves based on the past history of innovation levels in all regions and diffusion speeds in all region-pairs, capturing both the direct and indirect effects of the ICT shock. In each region and sector, a unit continuum of immobile firms hire inventors, own their ideas, and engage in Bertrand competition à la Bernard et al. (2003). The lowest-cost producer of each good captures the entire market for that good by charging the highest markup that deters competitors from entering. As a result, the expected value of an idea – and consequently the wages of innovation workers and incentives to innovate in each region – depends on equilibrium trade and idea diffusion shares. Over time, workers make dynamic decisions about moving between production and research and relocating across regions and sectors, based on relative wages in innovation and production, extending Caliendo et al. (2019). The resulting equilibrium worker mobility shares shape the spatial direction of innovation along the transition path, while 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, the 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. The equilibrium objects and central mechanisms of the model are presented as lemmas and propositions, with proofs relegated to Appendix C.

3.1 Endogenous Innovation and Technology Diffusion

Research in each region and sector is produced by firms, that hire local inventors (i.e. workers in the region-sector that choose innovation instead of production at time t^*), via the following production function:

$$\lambda_{r,t^*}^{k} = \underbrace{A_{r,t^*}}_{\substack{\text{fundamental}\\ \text{research}\\ \text{productivity}\\ \text{in region } r\\ \textbf{(function of}\\ \textbf{college ratio)}}} \cdot \underbrace{A_{t^*}^{k}}_{\substack{\text{sector-specific}\\ \text{national research}\\ \text{productivity}\\ \textbf{(Direct effect}\\ \textbf{of ICT)}}} \cdot \underbrace{\left(L_{r,t^*}^{R}\right)^{\alpha}}_{\substack{\text{agglomeration}\\ \text{economies}\\ \text{in innovation}\\ \text{including}\\ \textbf{spillovers}\\ \textbf{from ICT}\\ \textbf{to non-ICT}}} \cdot \underbrace{L_{r,t^*}^{k,R}}_{\substack{\text{number}\\ \text{of inventors}}} \cdot \underbrace{T_{r,t^*}^{k}}_{\substack{\text{sector-specific}\\ \text{technology level}\\ \text{in region } r}} . \tag{8}$$

The first term represents the fundamental research productivity in the region. It includes a time-invariant component proportional to the region's college ratio – reflecting the comparative advantage of high-skill cities in innovation – 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 common across all regions. This term captures the compositional shift of innovation towards ICT, reflecting the direct effect of the ICT shock on national research productivity in the ICT sector relative to other sectors. The third term represents agglomeration economies in innovation, including spillovers from ICT to non-ICT innovation (and vice versa) – the first indirect effect of the shock. The fourth term denotes the number of inventors in the region and sector, directly scaling the production of new ideas. The fifth term captures the level of technology in the region and sector. This term ensures the model delivers fully endogenous growth on the balanced growth path, as shown by Eaton and Kortum

(2024) in a setting without idea applicabilities[10].

While this deterministic research production function captures compositional shifts and innovation spillovers resulting from the ICT shock, additional structure is necessary to derive the goods productivity distribution and equilibrium trade shares in the Eaton-Kortum structure. An alternative interpretation of the research production function, incorporating stochastic microfoundations, is that each inventor in region r and sector k receives idea draws from a marked Poisson process with rate $A^k_{t^*} A_{r,t^*} \left(L^R_{r,t^*}\right)^\alpha T^k_{r,t^*}\ dt$, such that the aggregate arrival rate in the region and sector aligns with equation (8). Each point on this Poisson process corresponds to an individual idea and is characterized by three independent marks: (i) the good or variety ν to which it applies, drawn from the uniform distribution over [0,1]; (ii) its intrinsic quality q drawn from a Pareto distribution with cumulative distribution function:

$$H(q) = 1 - q^{-\theta}, \qquad q \geq 1, \quad \theta > 0 \tag{9}$$

where the parameter θ reflects the variability in idea quality, and; (iii) its applicability a, which is distinct across regions o and times after discovery ($t \geq t^*$).

The applicability component captures technology diffusion at the level of individual ideas and is instrumental for both my central theoretical results and the tractability of my overall framework. Concretely, it is represented by an $N \times 1$ vector, where each element corresponds to a distinct region. This vector governs two key aspects for a given idea in each region: (i) the probability the idea diffuses to a given region at each $t \geq t^*$, and (ii) the suitability of the idea for producing the corresponding good in the region upon its diffusion there. In the idea's discovery region r, the vector's element is a deterministic path of 1 for all $t \geq t^*$, reflecting that the idea is always accessible in the location where it was initially developed. At the idea's discovery time, its applicability is drawn from the following Pareto distribution:

$$F(a) = 1 - \left(\frac{\overline{a}}{a}\right)^\sigma, \qquad a \geq 1, \quad \sigma > \theta, \quad \overline{a} = \Gamma\left(1 - \frac{\theta}{\sigma}\right)^{-\frac{1}{\theta}}. \tag{10}$$

In each of the other regions $o \neq r$, the idea arrival probability is stochastic. Specifically, the applicability vector's element is an independent marked Poisson process with the arrival rate $\delta_{ro,t^*}\ dt$ for all $t \geq t^*$, where the marks reflect the idea's suitability for local production of the corresponding good and are drawn from equation (10)[11]. This stochastic nature of applicability ensures that even high-quality ideas may have varying levels of usefulness across regions and periods, generating non-trivial equilibrium idea diffusion shares[12].

[10] 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).

[11] Note that this intuitive microfounded structure implies that the history of all the applications in region o of an idea developed in region r consists of points of a two-dimensional Poisson process with intensity $\Gamma\left(1 - \frac{\theta}{\sigma}\right)^{-\frac{\sigma}{\theta}} \sigma a^{-\sigma-1} \delta_{ro,t^*} e^{-\delta_{ro,t^*}(t-t^*)} dadt$, the formulation used in Lind and Ramondo (2024). That idea diffusion follows a Poisson process implies exponential diffusion as in Eaton and Kortum (1999), but with a random component so that the set of countries where the idea has diffused to need not be tracked. The scale parameter $\overline{a}$ in equation (10) ensures that the derived goods productivity distribution is finite.

[12] In the next subsection, I leverage these idea diffusion shares to provide the first analytical characterization of the degree of colocation of innovation and production.

3.2 Production and Trade

Given sector-specific innovation rates, the process of technology diffusion described above operates independently across sectors. Consequently, the productivity distributions and trade shares are independently determined for each sector. To enhance clarity, I omit sector superscripts in what follows.

At each time t and in each sector k, region o produces each good or variety ν using the most efficient idea i available to them:

$$z_{o,t}(\nu) = \max_{i} \{q_i \cdot a_{i,o,t}\} \tag{11}$$

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

This microfounded structure for innovation and technology diffusion is particularly tractable because it endogenously generates a multivariate Fréchet distribution for goods productivity. Formally:

Lemma 1. *Given that ideas arrive from a marked Poisson process described in equations (8-(10), each idea is discovered at time t^* at a unique discovery location, and its productivity is multiplicative in quality and applicability [equation 11], the* ***joint productivity distribution*** *across regions at each time t is max-stable multivariate Fréchet, given by:*

$$\mathbb{P}\left[Z_{1,t} \leq z_1, \ldots, Z_{N,t} \leq z_N\right] = \exp\left[-\sum_{o=1}^{N} \int_{-\infty}^{t} \left[\sum_{l=1}^{N} \Omega_{ro,t^*}(t-t^*)\ z_o^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} \lambda_{r,t^*} dt^*\right] \tag{12}$$

and the ***marginal productivity distribution*** *in each region is Fréchet and given by:*

$$\mathbb{P}\left[Z_{o,t} \leq z_o\right] = \exp\left[-T_{o,t} z_o^{-\theta}\right] \tag{13}$$

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

$$T_{o,t} = \sum_{r=1}^{N} T_{ro,t} = \sum_{r=1}^{N} \int_{-\infty}^{t} \underbrace{\Omega_{ro,t^*}(t-t^*)^{1-\rho}}_{\text{technology diffusion}} \cdot \underbrace{\lambda_{r,t^*}}_{\text{innovation}} dt^* \tag{14}$$

where $\Omega_{ro,t^*}(t-t^*)$[13] is the share of ideas discovered in region r at time t^* that arrived in region o by time t and is given by:

$$\Omega_{ro,t^*}(t-t^*) = \begin{cases} 1 - e^{-\delta_{ro,t^*}(t-t^*)} & \text{for } o \neq r, t \geq t^* \\ 1 & \text{for } o = r, t \geq t^* \\ 0 & \text{for } t < t^*. \end{cases} \tag{15}$$

[13]Note that I use the notation $\Omega_{ro,t^*}(t-t^*)$, following Lind and Ramondo (2024), to emphasize that the share of ideas that has diffused from region l to region r depends on the time from idea discovery $t-t^*$. In the next subsection, I leverage the dynamics of technology diffusion to provide the first formal characterization of the asymmetric scale effect: that high-skill cities benefit disproportionately from uniform increases in bilateral diffusion costs due to greater increases in idea market access.

In each region o, the production of each variety ν uses only labor with productivity $z_{o,t}(\nu)$ drawn from the equilibrium multivariate Fréchet distribution given by equation (12). A representative final goods producer aggregates across all varieties in each region and sector as follows:

$$Y_{o,t} = \left[\int_0^1 Y_{o,t}(\nu)^{\frac{\epsilon-1}{\epsilon}} \, d\nu\right]^{\frac{\epsilon}{\epsilon-1}}. \tag{16}$$

Trade costs are of the standard iceberg type, such that delivering one unit of any variety from region o to region d at time t requires shipping $\tau_{od,t} \geq 1$ units of the variety, 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'. The cost of each variety ν in region d is the minimum unit cost across all regions:

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

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

Lemma 2. *Given the multivariate Fréchet distribution of goods productivity across regions in equation* (12)*, equilibrium* ***trade shares*** *are given by:*

$$\pi_{od,t} = \sum_{r=1}^{N} \pi_{rod,t} = \sum_{r=1}^{N} \int_{-\infty}^{t} \underbrace{\varphi_{rod|rd,t^*t}}_{\substack{\textit{share of goods in } d \\ \textit{produced in } o \textit{ at } t \\ \textit{given ideas from } r \textit{ at } t^* \\ \textbf{\textit{(conditional idea}} \\ \textbf{\textit{adoption shares)}}}} \cdot \underbrace{\phi_{rd,t^*t}}_{\substack{\textit{share of goods in } d \textit{ at } t \\ \textit{using ideas from } r \textit{ at } t^* \\ \textbf{\textit{(idea market shares)}}}} dt^* \tag{17}$$

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

$$\varphi_{rod|rd,t^*t} = \frac{\left[1 - e^{-\delta_{ro}(t-t^*)}\right] (w_{o,t}\tau_{od,t})^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N}\left[1 - e^{-\delta_{ro'}(t-t^*)}\right] \left(w_{o,t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}}, \tag{18}$$

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

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

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

$$\Phi_{rd,t^*t} \equiv \sum_o \left(1 - e^{-\delta_{ro,t^*}(t-t^*)}\right)(w_{o,t}\tau_{od,t})^{-\frac{\theta}{1-\rho}} = \sum_o \varphi_{rod|rd,t^*t}. \tag{20}$$

The ***price index*** *in region* d *is:*

$$P_{d,t} = \Gamma\left[\sum_{r'=1}^{N}\int_{-\infty}^{t} \Phi_{r'd,t^*t}^{1-\rho}\lambda_{r',t^*}dt^*\right]^{-\frac{1}{\theta}} \tag{21}$$

where Γ *is a constant, and* $1 + \theta > \epsilon$ *guarantees a well-defined price index.*

The equilibrium trade shares in equation (17) illustrate the interconnections between innovation, diffusion, and trade. Unlike standard trade models, this framework allows for innovation and production to occur in different locations. Consequently, trade shares are represented as the sum of trilateral shares, π_{rod}, 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 developed 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 $\varphi_{rod|rd,t^*t}$ that region o is the lowest-cost location for producing these goods (conditional idea adoption shares).

3.2.1 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 (18) across different destination markets d, we obtain the (unconditional) idea adoption shares:

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

This expression represents the share of successful ideas developed 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 derive the first analytical characterization of the degree of colocation between innovation and production, as formalized in the following corollary:

Corollary 1. *For goods sold at time t using ideas developed 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:*

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

This ratio captures the relative share of goods produced in the idea's origin 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 that diffuse from r 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).

3.2.2 The Asymmetric Scale Effect of Rising Technology Diffusion Speeds

Additionally, the idea market shares (equation 19) incorporate the idea market access term (equation 20), enabling the first analytical characterization of the **asymmetric scale effect**, where high-skill cities benefit disproportionately from a uniform increase in bilateral diffusion speeds. This is formalized in the following corollary:

Corollary 2. *If bilateral diffusion speeds are symmetric ($\delta_{rr'} = \delta_{r'r}$) and trade costs are identical ($\tau_{rd} = \tau_{r'd}$), 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.:*

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

Intuitively, a symmetric increase in the bilateral diffusion speed disproportionately benefits the higher-wage region by improving its access to a lower-cost region for production, allowing it to better exploit cost efficiencies. In contrast, the lower-wage region benefits less, as its already competitive production costs limit the relative advantage of gaining access to a higher-cost region.

Together, these corollaries underscore the central endogenous outcomes of my theory – colocation and asymmetric scale effects – which align precisely with the spatial mechanics of innovation documented in my empirical findings. Moreover, my theory is highly tractable, relying solely on particular probabilistic assumptions on technology, like other models in the Eaton-Kortum world. Besides the constant Γ, the equilibrium trade shares and price index are independent of the specific market structure for goods production.

3.3 Market Structure and Incentives to Innovate

Modeling endogenous innovation, however, requires profits from production to generate incentives to innovate. Thus, I build on seminal papers that depart from perfect competition (Eaton and Kortum, 2001; Bernard et al., 2003). In each region and sector, a unit continuum of immobile firms hire inventors 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.

Proposition 1. *Given Bertrand competition, the* ***markup distribution*** *is invariant across idea discovery time t^*, production time t, idea discovery location r, origin o, destination d, and sector k and distributed Pareto:*

$$\mathbb{P}\left(M \leq m | M \geq 1\right) = \frac{b_{rd}(1,t^*,t) - b_{rd}(m,t^*,t)}{b_{rd}(1,t^*,t)} = \frac{b_{rod}(1,t^*,t) - b_{rod}(m,t^*,t)}{b_{rod}(1,t^*,t)} = 1 - m^{-\theta} = H(m), \quad (23)$$

where $b_{ld}(m,t^*,t)$ is the probability that an idea discovered in region r at time t^* will undercut the lowest cost competitor in region d at time t by m, and $b_{rod}(m,t^*,t)$ is the same probability but with an additional condition that production of the good occurs in region o.

Proposition 1 extends the result in Eaton and Kortum (2001) to a setting with technology diffusion, where the location and time of goods production can differ from those of idea production[14]. Intuitively, this generalization is possible because the distribution of idea applicability is independent of idea quality. 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.

Specifically, the invariant markup distribution implies that all firms in sector k selling in destination d charge a markup drawn from $H(m)$. Consequently, the total profits earned from selling in destination d, irrespective of the locations of innovation and production, are given by:

$$\Pi_{d,t} = X_{d,t} \int_0^1 1 - \frac{1}{m(\nu)} d\nu = X_{d,t} \int_1^\infty 1 - \frac{1}{m(\nu)} dH(m) = \frac{X_{d,t}}{1+\theta} \tag{24}$$

where $X_{d,t}$ is the total spending by production workers and inventors in region d at time t and given by equation (35). Because the markup distribution is identical whether conditional or unconditional on the production location, profits from selling can be arbitrarily assigned to production, innovation, or a combination of both. For simplicity, I assume that all profits are allocated to innovation. With these profit transfers in each period, the expected value of an idea is formalized in the following lemma:

Lemma 3. *Given Bertrand competition and the time- and region-invariant markup distribution, the **expected value of an idea** in region r and sector k is:*

$$\check{V}_{r,t^*} = \int_{t^*}^\infty e^{-\zeta(t-t^*)} \sum_{d=1}^N \underbrace{\frac{\phi_{rd,t^*t}}{\lambda_{r,t^*}}}_{\substack{\text{share of profits earned}\\ \text{in region } d \text{ at time } t\\ \text{by an idea discovered}\\ \text{in region } r \text{ at time } t^*}} \cdot \underbrace{\frac{X_{d,t}}{1+\theta}}_{\substack{\text{profits earned}\\ \text{in region } d\\ \text{at time } t\\ \text{by all ideas}}} \cdot \underbrace{\frac{P_{rt^*}}{P_{rt}}}_{\substack{\text{accounting for}\\ \text{changes in}\\ \text{purchasing power}\\ \text{over time}}} dt \tag{25}$$

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 presented 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 prices over time.

Firms generate a stream of profits from the ideas they own, which they reinvest continuously in risk-free assets. Assets are produced using the same technology as final goods in each region[15], as described by equation (16), and do not depreciate over time. At each time t^* and in each region-sector, firms hire local inventors to produce ideas on their behalf, compensating them with wages equal to the expected return from

[14] Eaton and Kortum (2024) provide a more general derivation of the markup distribution by leveraging the distribution of price gaps, but do not address how this distribution might change in a setting with technology diffusion.

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

their innovation efforts:

$$w_{r,t^*}^{k,R} = A_{t^*}^{k} A_{r,t^*} \left(L_{r,t^*}^{R}\right)^{\alpha} \check{V}_{r,t^*}^{k}. \tag{26}$$

Thus, the wages of inventors are simply the product of the expected number of ideas they produce, as described by the idea production function in equation (8), and the expected value of each idea.

The ratio of sector-specific inventor real wages across regions, in turn, provides key insights into the spatial direction of innovation, as formalized in the following proposition:

Proposition 2. *Given Bertrand competition and the structure of inventor compensation by firms, the* ***spatial direction of innovation*** *is governed by the ratio of sector-specific inventor real wages across regions:*

$$\frac{\omega_{r,t^*}^{k,R}}{\omega_{r',t^*}^{k,R}} = \underbrace{\frac{A_{r,t^*}}{A_{r',t^*}}}_{\substack{\textit{fundamental} \\ \textit{research} \\ \textit{productivity} \\ \textit{(function of} \\ \textbf{\textit{college ratio)}}}} \cdot \underbrace{\left(\frac{L_{r,t^*}^{R}}{L_{r',t^*}^{R}}\right)^{\alpha}}_{\substack{\textit{agglomeration} \\ \textit{economies} \\ \textit{in innovation} \\ \textit{including } \textbf{\textit{spillovers}} \\ \textbf{\textit{from ICT to non-ICT}}}} \cdot \underbrace{\frac{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\boldsymbol{\phi_{rd,t^*t}}}{\boldsymbol{\lambda_{r,t^*}}} \frac{Y_{d,t}}{1+\theta} \frac{1}{P_{r,t}} dt}{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\boldsymbol{\phi_{r'd,t^*t}}}{\boldsymbol{\lambda_{r',t^*}}} \frac{Y_{d,t}}{1+\theta} \frac{1}{P_{r',t}} dt}}_{\substack{\textit{expected market potential of an idea,} \\ \textit{with the } \textbf{\textit{idea market shares}} \textit{ capturing} \\ \textbf{\textit{colocation between innovation and}} \\ \textbf{\textit{production}} \textit{ and } \textbf{\textit{idea market access}}}} \tag{27}$$

The higher the ratio in equation (27), the greater the returns to innovation directed toward region r relative to region r'. Within each sector, the incentives to innovate across space are shaped by three main factors: (i) time-varying fundamental research productivity, including a time-invariant component tied to regional college ratios; (ii) agglomeration economies in innovation, and; (iii) the expected market potential of an idea in all future periods. The first term captures forces unexplained by the model, such as local policies, infrastructure, or other region-specific factors, as well as the comparative advantage of high-skill regions in innovation. The second term reflects the benefits of proximity, including local knowledge spillovers from ICT to non-ICT innovation – capturing the first indirect effect of the ICT shock as documented in my empirical findings. The third term illustrates how idea market shares directly influence incentives to innovate. As demonstrated in the previous subsection, these equilibrium idea market shares incorporate the idea market access term, which capture the asymmetric scale effect – the second indirect effect of the ICT shock. This idea market access term is composed of the conditional idea adoption shares, which reflect the degree of colocation of innovation and production – the mechanism driving the direct effect of the ICT shock.

This ratio serves as an analog to the sectoral direction of technical change in Acemoglu (1998, 2002, 2007), capturing inventor real wages in partial equilibrium to highlight the main forces shaping the spatial direction of innovation. It is important to note that inventor real wages are determined by the market clearing condition in general equilibrium and hence influenced by many other factors in the economy, such as production worker wages in all regions. In Appendix C.7, I present the wage ratio that governs the sectoral direction of innovation.

3.4 Consumption and Worker Mobility

These incentives to innovate shape the spatial distribution of innovation rates because, as in any model of endogenous innovation, workers must decide between production and research. To account for the observed gradual rise in the spatial concentration of innovation following the ICT shock, my theory incorporates bilateral mobility frictions. Additionally, to reflect the underlying trend of workers migrating to high-skill cities, workers can move not only between production and research but also across regions and sectors. This dynamic worker mobility decision extends the framework in Caliendo et al. (2019). These bilateral decisions necessitate additional notation. I denote worker occupations h, n to represent either production (G) or research (R).

At each time t, given a worker's choice of region d, sector k, and occupation h, they supply a unit of labor inelastically, receive wages, and consume a bundle of local final goods from the ICT and non-ICT sectors with the following Cobb-Douglas preferences:

$$C_{d,t}^{k,h} = \left(c_{d,t}^{k,h,\text{ICT}}\right)^{\iota} \left(c_{d,t}^{k,h,\text{non-ICT}}\right)^{1-\iota}. \tag{28}$$

where ι captures the sectoral expenditure share allocated to ICT goods.

Over time, a Poisson arrival process with rate 1 governs when 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:

$$v_{d,t}^{k,h} = \max_{o,s,n} \quad \frac{w_{d,t}^{k,h}}{P_{d,t}} + \frac{1}{1+\zeta}\mathbb{E}_t\left(\mathbb{E}_{\epsilon}\left[v_{o,t'}^{s,n}\right]\right) - \kappa_{do,t}^{ks,hn} + \epsilon_{o,t}^{s,n} \tag{29}$$

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:

$$\breve{F}\left(\left\{\epsilon_{o,t}^{s,n}\right\}_{o=1,\ldots,N}^{s=\{\text{ICT},\text{non-ICT}\},n=\{G,R\}}\right) = \exp\left\{-\left[\sum_{o}\sum_{s}\left(\sum_{n}\exp\left(-\epsilon_{o,t}^{s,n}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right]\right\} \tag{30}$$

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 4. *Given individual-level worker mobility decisions defined by equations* (29)-(30)*, the* ***expected value*** *or lifetime utility of a representative worker in labor market* (d, k, h) *is given by:*

$$V_{d,t}^{k,h} \equiv \mathbb{E}_{\epsilon}\left[v_{d,t}^{k,h}\right] = \frac{w_{d,t}^{k,h}}{P_{d,t}} + \frac{1}{\Upsilon}\log\left[\sum_{o}\sum_{s}\left(\sum_{n}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right] \tag{31}$$

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

$$\begin{aligned}\mu_{do,t}^{ks,hn} &\equiv \mathbb{E}_t\left[\breve{\mu}_{do,t'}^{ks,hn}\right] \equiv \mu_{do,t}^{ks,hn}|\mu_{do,t}^{ks} \cdot \mu_{do,t}^{ks} \\ &= \underbrace{\frac{\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}}}_{\text{switching between production and research}} \cdot \underbrace{\frac{\left[\sum_{n'}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{o'}\sum_{s'}\left[\sum_{n'}\exp\left(\frac{1}{1+\zeta}V_{o',t'}^{s',n'} - \kappa_{do',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}}_{\text{migration across regions and sectors}}\end{aligned} \tag{32}$$

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

$$L_{o,t'}^{s,n} = \sum_{h}\sum_{k}\sum_{d}\mu_{do,t}^{ks,hn}L_{d,t}^{k,h}. \tag{33}$$

3.5 Market Clearing

Closing the economy requires a market clearing condition for each time t, as formalized in the following lemma:

Lemma 5. *Accounting for the transfer of profits across regions, the combined goods and innovation* ***market clearing*** *condition at time t is given by:*

$$\frac{1+\theta}{\theta}w_{o,t}^{k}L_{o,t}^{k} = \sum_{d}\pi_{od,t}^{k}\iota^{k}\left[\sum_{k}\left(w_{d,t}^{k}L_{d,t}^{k} + \sum_{r}\varphi_{dr,t}^{k}\frac{1}{\theta}w_{r,t}^{k}L_{r,t}^{k}\right)\right] \tag{34}$$

where $\pi_{od,t}^{k}$ *are the trade shares from equation* (17) *and* $\varphi_{dr,t}^{k}$ *are the idea adoption shares from equation* (22)*.*

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

$$X_{d,t}^{k} = \iota^{k}\left[\sum_{s}\left(w_{d,t}^{s}L_{d,t}^{s} + \sum_{r}\varphi_{dr,t}^{s}\frac{1}{\theta}w_{r,t}^{s}L_{r,t}^{s}\right)\right]. \tag{35}$$

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 have omitted the superscript G for wages and labor to reduce notational burden.

3.6 Definition of Equilibrium

Given an initial distribution of technology levels $\{T_{o,0}^{k}\}_{o=1,k=1}^{N,N}$ and workers $\{L_{o,0}^{k,h}\}_{o=1;k;h}^{N;\{\text{ICT,non-ICT}\};\{G,R\}}$, trajectories of bilateral trade costs $\{\tau_{od,t}\}_{o=1,d=1,t=0}^{N,N,\infty}$, bilateral migration costs $\{\kappa_{od,t}^{ks,ha}\}_{o=1;d=1;t=0;k,s;h,a}^{N;N;\infty;\{\text{ICT,non-ICT}\};\{G,R\}}$, bilateral diffusion lags $\{\delta_{od,t^*}\}_{o=1,d=1,t^*=0}^{N,N,\infty}$, fundamental productivities in idea production $\{A_{r,t^*}, A_{t^*}^{k}\}_{r=1,k,t^*=0}^{N,\{\text{ICT,non-ICT}\},\infty}$ 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 4), evolution of technology levels (Lemma 1 equation 14), bilateral trade and idea adoption shares (Lemma 2 equations 17 and 22), returns to innovation (Lemma 3) and market clearing condition (Lemma 5).

3.7 Model Extensions

By microfounding innovation and technology diffusion within the Eaton-Kortum framework, my dynamic spatial model 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

Beyond formalizing the key mechanisms driving the rising spatial concentration of innovation, my model provides a framework to analyze its aggregate consequences, particularly its impact on overall economic growth and the welfare implications of the ICT shock.

4.1 Balanced Growth Path

Specifically, the balanced growth path is highly flexible due to its block recursive nature, illustrating the causes and consequences of the geography of innovation. This structure is formalized in the following proposition:

Proposition 3. *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:

$$\dot{T}_o^k(t) = \sum_r \gamma_r^k T_r^k(t) \int_{-\infty}^{t} g^k e^{-g^k(t-t^*)} \Omega_{ro}(t-t^*)^{1-\rho} dt^* \tag{36}$$

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

$$g\mathbf{T}^k = \mathbf{\Delta}^k(g)\mathbf{T}^k \tag{37}$$

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:

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

Thus, g^k is the Perron-Frobenius root of equation (37) 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 $g_v = \frac{1+\zeta}{\zeta}\frac{1}{\theta}\sum_k \iota^k g^k$.

The first part of the proposition demonstrates how arbitrary and heterogeneous regional innovation rates ($\gamma_r^k = \frac{T_r^k}{\lambda_r^k}$) and region-pair idea diffusion speeds (δ_{ro}) deliver 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. Notably, reduced communication costs resulting from the ICT shock directly impact aggregate growth, in addition shaping the geography of innovation. More broadly, since any spatial distribution of innovation rates is consistent with a balanced growth path, my model can flexibly accommodate a wide range of alternative mechanisms that influence the geography of innovation in other contexts.

The second part of the proposition describes how the remaining variables in the economy evolve along the balanced growth path. Specifically, the growth rate of the expected value of workers, which captures the aggregate growth rate of the economy, is an explicit function of the growth rates of technology in different sectors, while the steady-state distribution of workers determines heterogeneous innovation rates across regions and sectors. This steady-state distribution of workers, 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.8.

4.2 Transition Path

Beyond balanced growth, I characterize the transition path to illustrate how the spatial distribution of innovation and the overall economy gradually evolved following the ICT shock in the early 1990s. To align my model with observed data on innovation levels and trade shares, which are available at most at an annual frequency, I introduce additional assumptions that govern the timing of innovation and production. Specifically, while inventors receive ideas from a Poisson process described in Section 3.1, a separate Poisson arrival process with rate 1 determines when inventors can produce their ideas. Additionally, an independent Poisson arrival process with rate 1 governs when firms can produce goods. These Poisson processes ensure that production and innovation occur at discrete moments in time, aligning with the frequency of observed innovation levels and trade shares, even though the model operates in continuous time. Let $\widetilde{x}_t = x_t e^{-g_x t}$ denote the detrended value, $\widehat{x}_{t'} = \frac{\widetilde{x}_{t'}}{\widetilde{x}_t}$ represent changes in the detrended value for any variable x with growth rate g_x, and $u = \exp(V)$. The transition path can then be fully characterized by 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:

Proposition 4. *Given an initial distribution of workers, wages, technology levels, and migration and trade shares* $\left\{L_{o,0}^{k}, T_{o,0}^{k}, \mu_{od,0}^{ks,hn}, \pi_{od,0}^{k}\right\}_{o=1;d=1;k,s;h,n}^{N;N;\{ICT,non\text{-}ICT\};\{G,R\}}$*, exogenous trajectories of changes in migration and trade costs, research productivity, and bilateral diffusion lags* $\left\{\dot{\kappa}_{od,t}^{ks,hn}, \dot{\tau}_{od,t}, A_{r,t}^{k}, \delta_{ro,t}\right\}_{o=1;d=1;r=1;k,s;h,n;t=0}^{N;N;N;\{ICT,non\text{-}ICT\};\{G,R\};\infty}$, *and additional Poisson processes that govern the timing of innovation and production, the* ***transition path*** *of the economy is characterized by the evolution of the distribution of workers for all* $t^*, t, t' \leq T$*:*

$$\log\left(\widehat{u}_{d,t}^{k,h}\right) = \log\left(\frac{\widehat{w}_{d,t}^{k,h}}{\widehat{P}_{d,t}}\right) + \frac{1}{\Upsilon}\log\left[\sum_{n}\left[\sum_{s}\sum_{o}\mu_{do,t'}^{ks,hn}\left(\widehat{u}_{o,t'}^{s,n}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}\right] \tag{38}$$

$$\mu_{od,t'}^{ks,hn} = \frac{\mu_{od,t}^{ks,hn}\left(\widehat{u}_{d,t'}^{s,n}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'}\mu_{od,t}^{ks,hn'}\left(\widehat{u}_{d,t'}^{s,n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}}\cdot\frac{\left[\sum_{n'}\mu_{od,t}^{ks,hn'}\left(\widehat{u}_{d,t'}^{s,n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{n'}\left[\sum_{d'}\sum_{s'}\mu_{od',t}^{ks',hn'}\left(\widehat{u}_{d',t'}^{s',n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}} \tag{39}$$

$$L_{d,t'}^{k,h} = \sum_{n}\sum_{s}\sum_{o}\mu_{od,t}^{ks,hn}L_{o,t}^{s,n} \tag{40}$$

where at each time t *innovation levels are given by equation* (8) *and technology levels by:*

$$\begin{aligned}\lambda_{r,t}^{k} &= A_{t^*}^{k}A_{r,t^*}\left(L_{r,t^*}^{R}\right)^{\alpha}L_{r,t^*}^{k,R}T_{r,t^*}^{k}\\ T_{o,t}^{k} &= \sum_{r=1}^{N}\sum_{t^*\in\mathfrak{T}_0^t}\Omega_{ro,t^*}(t-t^*)^{1-\rho}\cdot\lambda_{r,t^*}^{k} + T_{o,0}^{k},\end{aligned} \tag{41}$$

with $\mathfrak{T}_0^t$ *as the set of innovation times* t^* *from time* 0 *to time* t*, the trade equilibrium is given by:*

$$\pi_{od,t}^{k} = \sum_{r=1}^{N}\sum_{t^*\in\mathfrak{T}_0^t}\frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}^{k,G}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{r,t}^{k,G}\tau_{ro',t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^*}^{k}}{\sum_{r'}\sum_{\check{t}\in\mathfrak{T}_0^t}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^{k}} \tag{42}$$

$$\varphi_{ro,t}^{k} = \sum_{d=1}^{N}\sum_{t^*\in\mathfrak{T}_0^t}\frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}^{k,G}\tau_{od,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{r,t}^{k,G}\tau_{ro',t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^*}^{k}}{\sum_{r'}\sum_{\check{t}\in\mathfrak{T}_0^t}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^{k}} \tag{43}$$

$$P_{d,t}^{k} = \Gamma\left[\sum_{r'=1}^{N}\sum_{t^*\in\mathfrak{T}}\left[\sum_{o'=1}^{N}\Omega_{r'o',t^*}(t-t^*)\left(w_{o',t}^{k,G}\tau_{o'd,t}^{k}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',t^*}^{k}\right]^{-\frac{1}{\theta}} \tag{44}$$

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

where the market clearing condition comes from equation (34)*, 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:

$$w_{r,t}^{k,R} = A_t^k A_{r,t} \left(L_{r,t}^{k,R}\right)^{\alpha} T_{r,t}^k \check{V}_{r,t}^k = \frac{\check{V}_{r,t}^k \lambda_{r,t}^k}{L_{r,t}^{k,R}} \tag{45}$$

$$\check{V}_{r,t}^k = \sum_{t' \in \mathcal{T}_t^{\infty}} \left(\frac{1}{1-\rho}\right)^{t'-t} \sum_{d=1}^{N} \frac{X_{d,t'}^k}{1+\theta} \cdot \frac{P_{r,t}}{P_{r,t'}} \cdot \frac{\phi_{rd,tt'}^k}{\lambda_{l,t}^k} \tag{46}$$

$$X_{d,t'}^k = \iota^k \left[\sum_k \left(w_{d,t'}^{k,G} L_{d,t'}^{k,G} + \sum_l \varphi_{dr,t'}^k \frac{1+\theta}{\theta} w_{r,t'}^{k,G} L_{r,t'}^{k,G}\right)\right] \tag{47}$$

$$\phi_{rd,tt'}^k = \frac{\left[\sum_{o'} \Omega_{ro',t}(t'-t) \left(w_{r,t'}^k \tau_{ro',t'}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} \lambda_{r,t}^k}{\sum_{r'} \sum_{\check{t} \in \mathcal{T}_{-\infty}^{t'}} \left[\sum_{o'} \Omega_{r'o',\check{t}}(t'-\check{t}) \left(w_{o',t'}^k \tau_{o'd,t'}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} \lambda_{r',\check{t}}^k} \tag{48}$$

and the growth rate of prices is given by $g_p = \frac{1}{\theta} \sum_k \iota^k g^k$ with g^k as the Perron-Frobenius root of equation (37)*, as described in Proposition 3.*

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 (39) and (40) – along with exogenous research productivities, determine the trajectory of innovation and technology levels given by equations (8) and (41). The trajectory of innovation levels, along with exogenous bilateral diffusion lags, determine the trade equilibrium for each production time, as characterized by equations (42)-(44) and (34). 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 (48). Trajectories of this probability determine incentives to innovate, captured by the value of individual ideas in equation (46) and hence inventor wages in equation (45). In turn, the wages of production workers and inventors determine the incentives of workers to migrate – given by equation (38) – and hence determine the evolution of the distribution of workers across regions, sectors, and occupations.

Notice that the trade equilibrium cannot be expressed in time differences. This is because trade shares depend on the entire trajectory of past innovations, rather than just the contemporaneous technology stock. Thus, in principle, information on levels of exogenous trade costs are required to obtain the transition path. However, given data on wages and trade shares in the initial period – where the economy is assumed to be in steady state and all prior ideas have fully diffused – initial trade costs can be obtained from equation (42), which collapses to the canonical Eaton-Kortum trade shares:

$$\pi_{od,0}^k = \frac{T_{o,0}^k \left(w_{o,0}^k \tau_{od,0}^k\right)^{-\theta}}{\sum_{o'=1}^{N} T_{o',0}^k \left(w_{o',0}^k \tau_{o'd,0}^k\right)^{-\theta}}.$$

Hence, simulating the transition path only requires changes in trade costs, because these changes can be converted to levels given initial trade costs. Note also that the trade equilibrium does not contain detrended variables. This is because production worker wages are constant on the balanced growth path and can be determined without the sectoral price index, which declines over time.

4.3 Welfare Impacts of the ICT Shock

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 $\ddot{x}_{t'} = \frac{\widehat{\acute{x}}_{t'}}{\widehat{\widetilde{x}}_{t'}} = \frac{\acute{\widetilde{x}}_{t'}/\acute{\widetilde{x}}_t}{\widetilde{x}_{t'}/\widetilde{x}_t}$ 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:

Corollary 3. *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:*

$$\log\left(\delta_d^{k,h}\right) = \sum_{t' \in \mathcal{T}_{\mathbb{R}+}} \left(\frac{1}{1+\zeta}\right)^{t'} \log\left(\underbrace{\frac{\ddot{w}_{d,t}^{k,h}}{\ddot{P}_{d,t}}}_{\substack{\text{change in}\\ \text{future}\\ \text{detrended}\\ \text{real wages}}} \underbrace{\frac{1}{\left(\ddot{\mu}_{dd,t}^{kk}\right)^{1/\Upsilon} \left(\ddot{\mu}_{dd,t}^{kk,hh}|\mu_{dd,t}^{kk}\right)^{\upsilon/\Upsilon}}}_{\text{change in option value of migration}} \right) + \underbrace{\frac{1}{\theta}\sum_k \iota^k \left(\acute{g}^k - g^k\right)}_{\text{growth effects}}. \quad (49)$$

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 relative to Caliendo et al. (2019). First, the option value of migration is given by both the own-migration share across region-sectors as well as the conditional own-migration share across occupations, since the idiosyncratic preferences of workers for each occupation is correlated across region-sectors. Second, the welfare expression includes impacts on long-run growth, since a counterfactual change in fundamentals has both transitory and long run effects in my model.

5 Conclusion

The emergence 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 challenging 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 rise 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 reduced communication costs enabled by ICT – 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 novel model of spatial growth that integrates endogenous innovation with technology diffusion at the level of individual ideas. In my model, the ICT shock is characterized by two exogenous components: the rising productivity of ICT innovation, which captures the compositional shift of innovation toward ICT, and falling bilateral diffusion speeds, reflecting reduced communication costs. The theory incorporates spillovers from ICT to non-ICT innovation through the idea production function and offers the first 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 benefits high-skill cities due to greater increases in idea market access. An additional advantage of my model is its ability to address the aggregate consequences of the ICT shock. Along the transition path, the ratio of real wages from innovation across regions captures the incentives for workers mobility, which in turn drives the spatial direction of innovation. Along the balanced growth path, prices in all regions fall at the same aggregate rate, endogenously determined by the steady state distribution of innovation rates. I then use these characterizations of the balanced growth and transition paths to analytically decompose the welfare impact of the ICT shock into its transitory and long-run growth components.

More broadly, my model integrates endogenous and directed innovation into existing dynamic spatial models in a highly tractable manner, offering methodological tools to explore the mechanics of innovation across space. 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.

Akcigit, U., Grigsby, J., Nicholas, T., and Stantcheva, S. (2021). Taxation and Innovation in the Twentieth Century. *The Quarterly Journal of Economics*, 137(1):329–385.

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

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. (1999). International technology diffusion: Theory and measurement. *International Economic Review*, 40(3):537–570.

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.

Eckert, F. and Peters, M. (2023). Spatial Structural Change. *Working Paper*.

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*.

Herrendorf, B., Rogerson, R., and Valentinyi, Á. (2014). *Growth and Structural Transformation*, volume 2. Elsevier B.V.

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.

Kingman (1992). *Poisson Processes*, volume 3. Clarendon Press.

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

Lind, N. and Ramondo, N. (2023b). Trade with Correlation. *American Economic Review*, 113(2):317–353.

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)[16]:

> *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[17]. 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.

[16]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.

[17]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[18].

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[19]. Other classifications are denoted as having a "later" position. USPTO Guideline 905.03(a)III.A.(C)[20] states that:

> *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[21]. I use the 08/02/2022 version, downloaded on 09/09/2022[22]. Just as Akcigit et al. (2021) use the primary USPC class of each patent, 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[23], 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**

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

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

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

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

[22]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 approximately 3–5% of patents for CPC classes and WIPO technology fields.

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

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 technology 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 13 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).

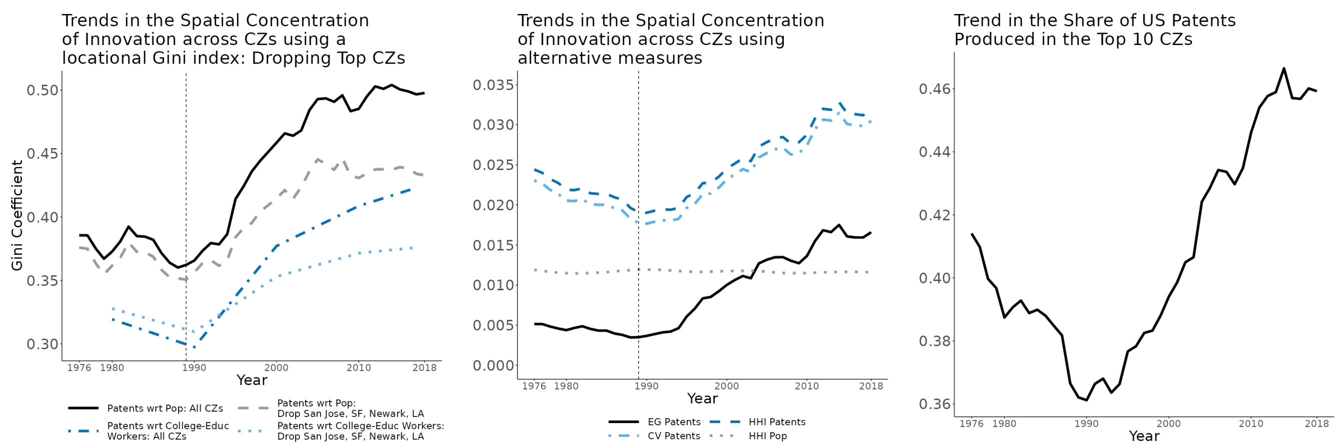


Figure 13: Robustness of trends in the spatial concentration of innovation across CZs. The top left graph plots trends in the locational Gini index of patents with respect to population and college-educated workers after removing top CZs in innovation: San Jose, San Francisco, Newark, and Los Angeles. 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 share per unit 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 14 shows that the geographical sorting of patenting intensity in high-skill cities from 1990 is not mechanically driven by an increase in the college ratio in these cities after 1990. The left graph plots

	CZ	Change in patent share per unit pop share 1990-2015	Patent share per unit pop share	Rank of patent share per unit pop 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 patent share per unit 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 patent share per unit population share in 2015 but did not experience the greatest increase from 1990-2015.

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 middle graph plots trends in the semi-elasticity of changes in the log college ratio relative to 1980 with respect to the 1980 college ratio. In other words, each estimate on the graph is the gradient of the correlation plot The semi-elasticity of 0.10 in 1990 tells us that a 10% increase in the 1980 college ratio across CZs is associated with a 1% greater increase in the college ratio between 1980-1990. That the semi-elasticity did not increase in 2000, 2010, or 2018 tells us that **worker sorting primarily occurred in the 1980s**. To the best of my knowledge, this trend provides first evidence on the dynamics of worker sorting; Moretti (2013) and Diamond (2016) focus on cross-sectional correlations in the CZ college ratio or percent in 1980 against its changes from 1980-2000. To provide an equivalent comparison, the right graph plots the semi-elasticity of changes in patent share per unit population share against the 1980 college ratio. This semi-elasticity was generally falling in the 1980s, but increased sharply after 1990. Together, these graphs tell us that the geographical sorting of patenting intensity in high-skill cities from 1990 is not mechanically driven by an increase in the college ratio in these cities after 1990. Rather, there was either a lagged effect of worker sorting, or an independent shock that interacted with the regional worker skill intensities in 1990 such that patents per capita became more concentrated in high-skill cities after 1990.

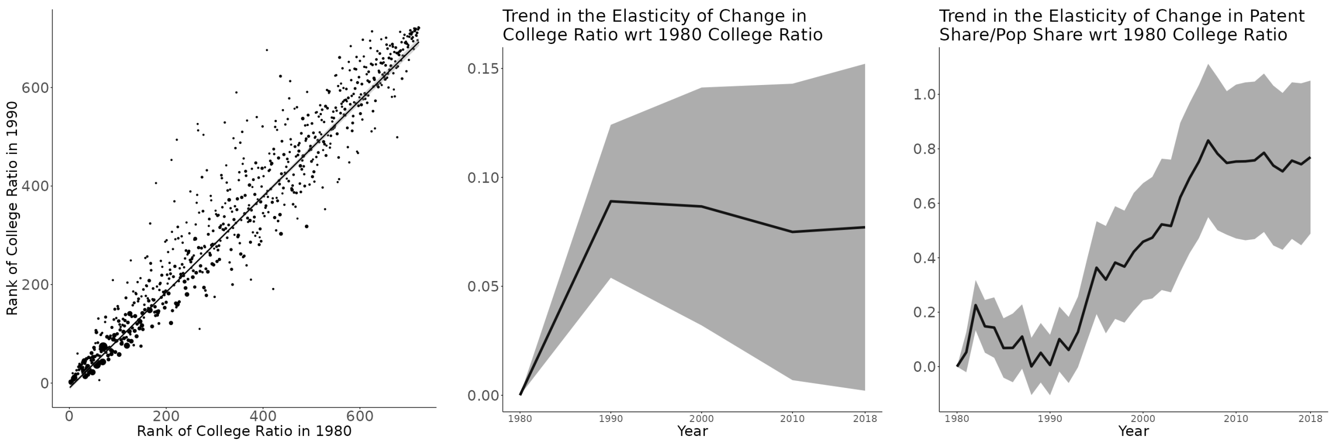


Figure 14: Trends in how percentage changes in patent share, population share, patent share per unit college share, and college ratio from 1990-2018 relate to the initial college ratio in 1990 across CZs, where each CZ is weighted by its population. The confidence intervals in reflect the statistical uncertainty of the slope estimates for each year given heteroskedastic-robust standard errors.

Figure 15 plots trends in the annual PPML 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 PPML 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 PPML elasticity of patents per capita with respect to the 1990 population density did not increase after 1990.

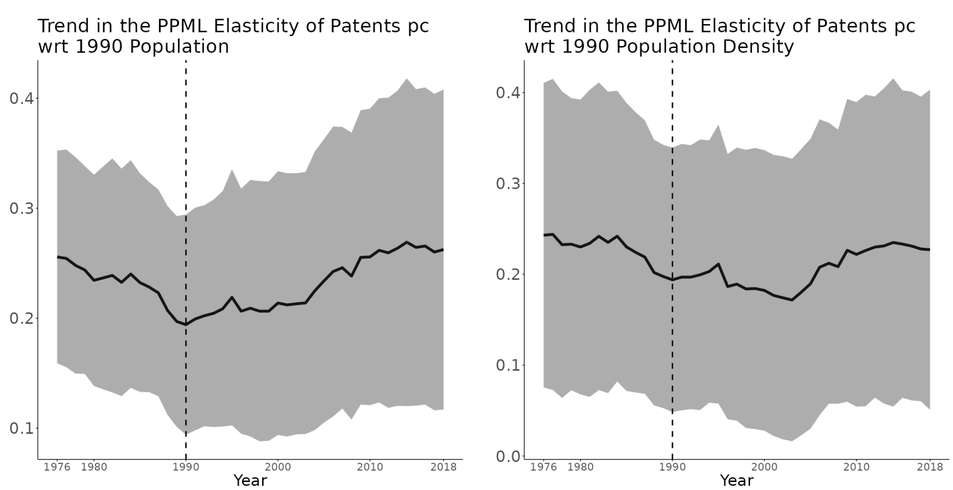


Figure 15: Trends in the annual PPML elasticity of CZ patents per capita with respect to the 1990 population (left) and 1990 population density (right).

B.2 Details of 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[24]

- 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.

[24]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 2 (Compositional Shift of Innovation towards ICT)

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:

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

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.

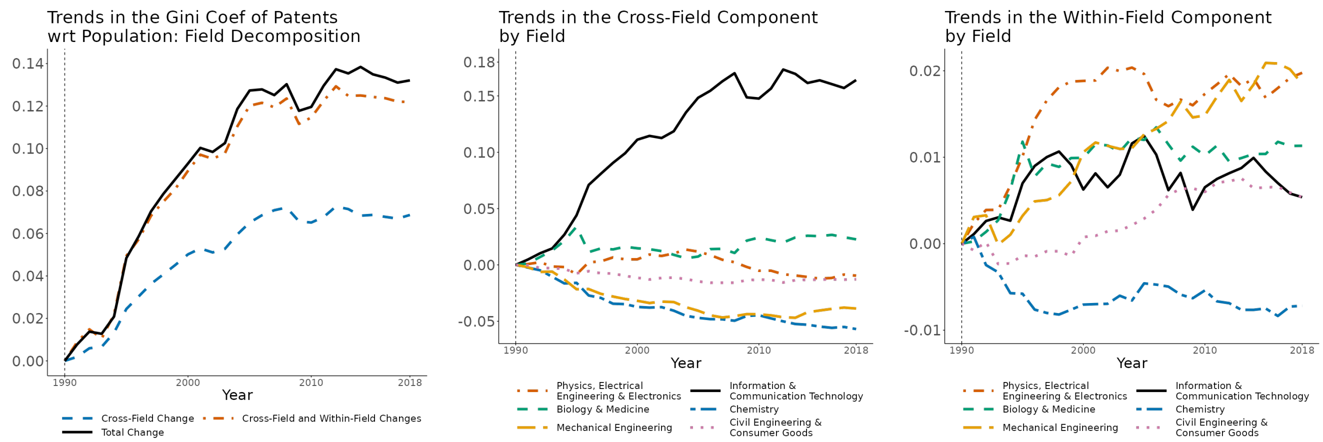


Figure 16: Trends in the decomposition of changes in the locational Gini coefficient of patents with respect to population relative to 1990. The left graph plots trends in the overall decomposition into within-field, across-field, and field colocation components while the middle and right graphs further decompose the cross-field and within-field components respectively into the contribution of each individual field.

The left graph in Figure 16 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 17 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.

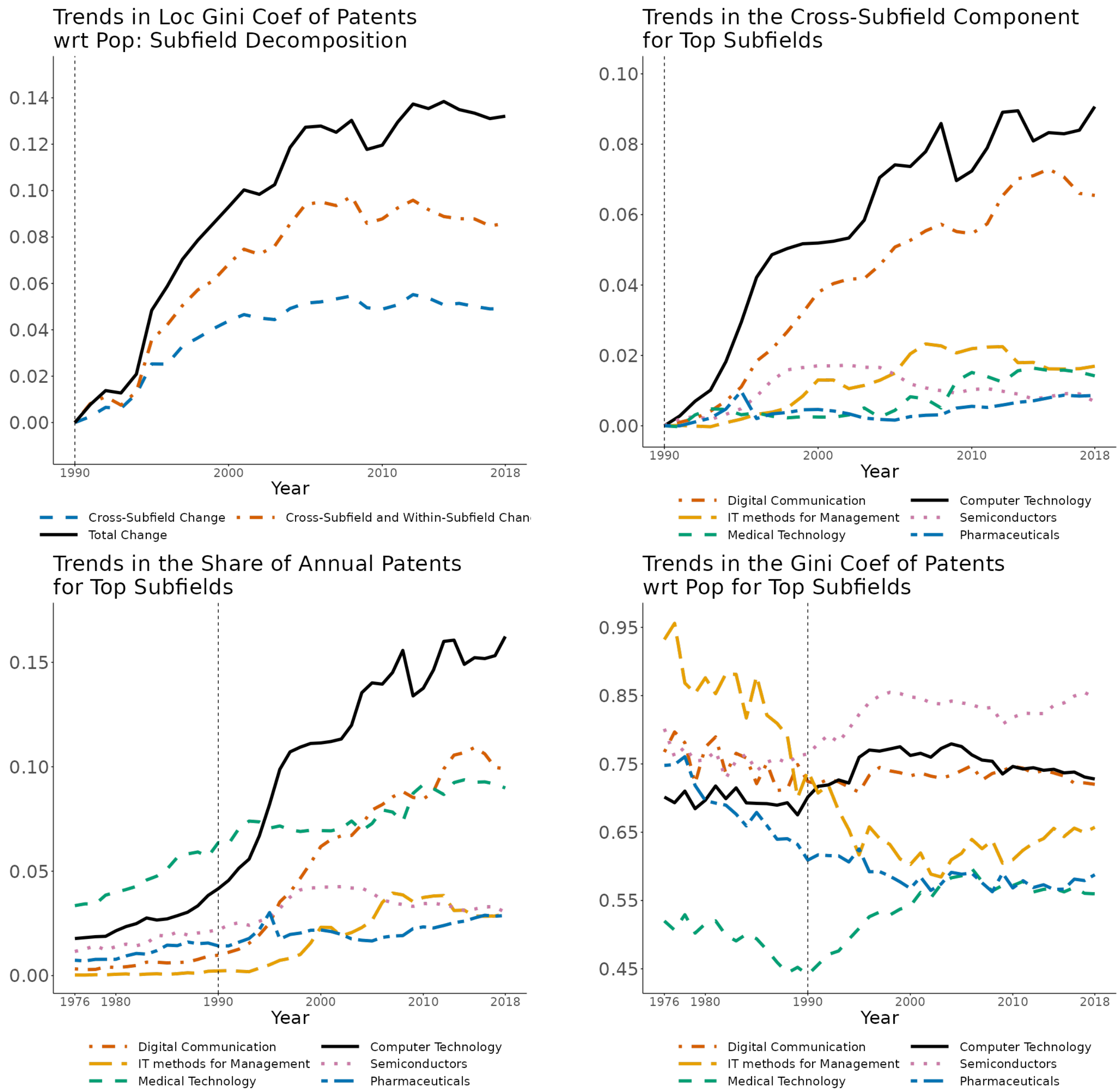


Figure 17: Trends in the subfield decomposition of changes in the aggregate locational Gini coefficient of patents with respect to population. The top left graph plots trends in the decomposition of the aggregate locational Gini coefficient into cross-subfield and within-subfield components. The top 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 18 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 6 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 16 – 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.

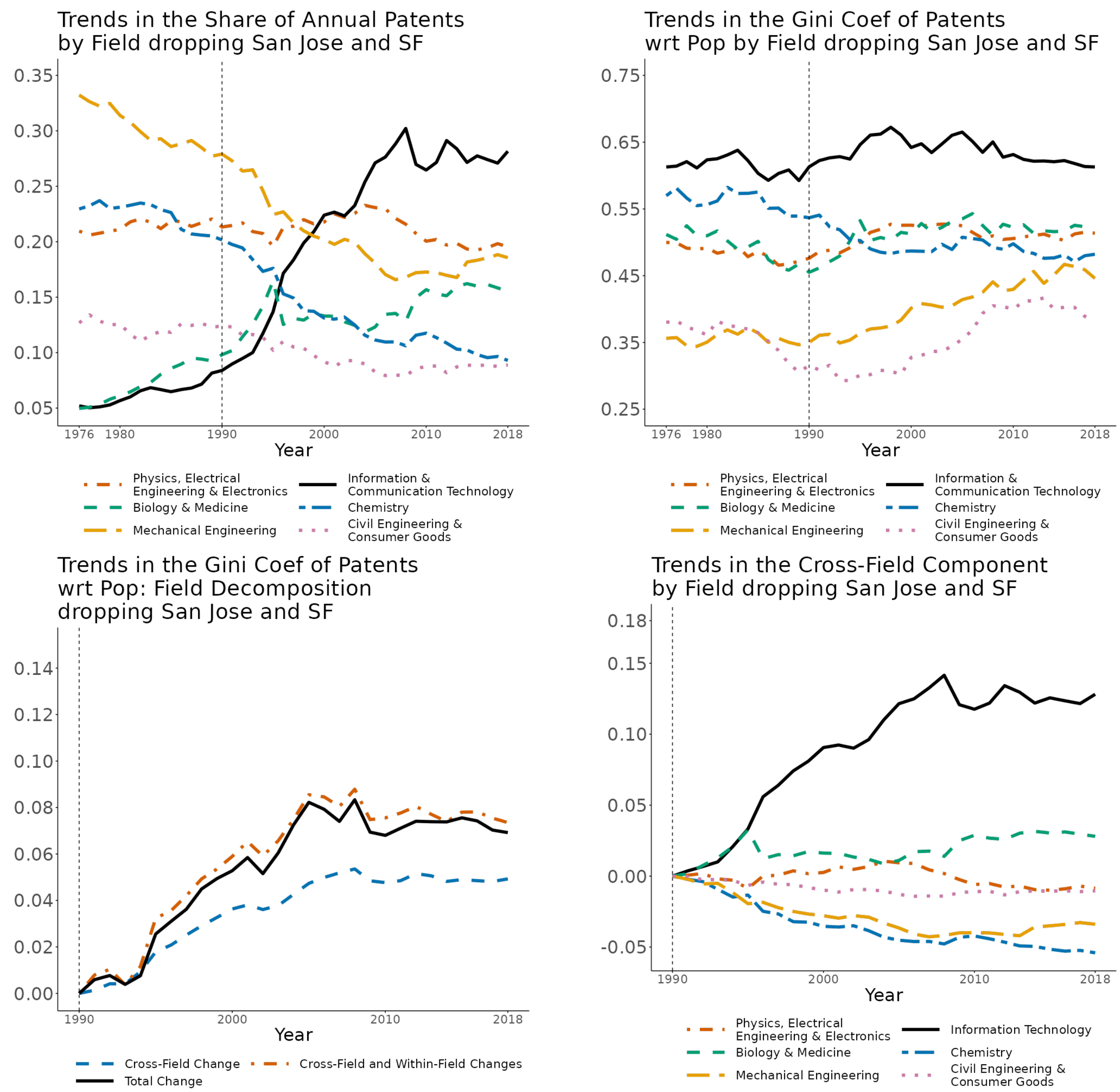


Figure 18: Trends in the share of annual patents by field (top left), the Gini coefficient of CZ patents with respect to CZ population by field (top right), decomposition of the overall Gini coefficient into within-field and cross-field components (bottom left), and a further decomposition of the cross-field component into each individual field (bottom right) after dropping San Jose and San Francisco.

C Proofs and Extensions of Propositions and Lemmas

C.1 Lemma 1 (Productivity Distribution)

Proof. My microfounded structure of innovation and technology diffusion, outlined in equations (8)-(10), is isomorphic to the independent multi-dimensional Poisson processes governing idea and applicability arrivals in Lind and Ramondo (2024). Since the goods productivity distribution is derived exclusively from the Poisson processes governing idea and applicability arrivals, my additions of endogenous innovation from Eaton and Kortum (2001) and worker mobility from Caliendo et al. (2019) do not change the expression or proof for the productivity distribution in Lind and Ramondo (2024) For more details on the multivariate Fréchet productivity distribution, see Lind and Ramondo (2023b). □

C.2 Lemma 2 (Trade Shares and Price Index)

Proof. See Lind and Ramondo (2024). Note that the trade shares and price index are derived exclusively from the multivariate productivity distribution. Thus, like in Lemma 1, my additions of endogenous innovation from Eaton and Kortum (2001) and worker mobility from Caliendo et al. (2019) to the model in Lind and Ramondo (2024) also do not change their expressions for equilibrium trade and prices. □

C.3 Lemma 3 (Expected Value of an Idea)

Proof. Total profits earned at time s in region d is given by $\Pi^k_{d,t} = \frac{X^k_{d,t}}{1+\theta}$, as described in equation 24. The share of these profits from ideas discovered in region r at time t^* is ϕ^k_{rd,t^*t}. Since the flow rate of ideas in region r at time t^* is λ^k_{r,t^*}, the expected flow of profits at time s in region d of an idea discovered in region r at time t^* is $\frac{\phi^k_{rd,t^*t}}{\lambda^k_{r,t^*}}\frac{X^k_{d,t}}{1+\theta}$. Accounting for changes in the purchasing power in region r over time and discounting future flows yields equation 25. Note that ϕ^k_{ld,t^*t} conditions on idea cohort and hence is a direct measure on whether the idea remains the lowest cost one in destination d, thereby allowing for a slightly simpler computation of value relative to Eaton and Kortum (2001). □

C.4 Lemma 4 (Expected Worker Value and Worker Mobility Shares)

Proof. Note that the individual worker mobility problem, as defined by equations 29-30, 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 32, 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 workers can move, with an arrival rate of 1, as described 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 29 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 30, with the correlation function defined as:

$$\check{F}\left(\left\{\exp\left(-\epsilon_{o,t}^{s,n}\right)\right\}_{o=1,\ldots,N}^{s=\{\text{ICT},\text{non-ICT}\},n=\{G,R\}}\right) = \sum_o \sum_s \left(\sum_n \exp\left(-\epsilon_{o,t}^{s,n}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}, \qquad (50)$$

where υ captures the correlation between production and research across region-sectors, and Υ is the scale parameter that adjusts the sensitivity of 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 (30). 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:

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

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:

$$\Phi_{d,t}^{k,h}=\frac{1}{\Upsilon}\log\left[\sum_{o,s}\left(\sum_{n}\exp\left(\frac{1}{1+\zeta}V_{o,t'}^{s,n}-\kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right],\tag{51}$$

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 29 yields the expected worker value in equation 31.

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 51 – is related to this conditional option value as follows:

$$\Phi_{d,t}^{k,h}=\frac{1}{\upsilon}\log\left[\sum_{o,s}\exp\left(\Phi_{do,t}^{ks,h}\right)^{\upsilon}\right].\tag{52}$$

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).

Thus, this proof 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.5 Lemma 5 (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 follow Eaton and Kortum (2024) and 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:

$$\begin{aligned} w_{o,t}^{k,G} L_{o,t}^{k,G} + \Pi_{o,t}^{k\;*} &= \sum_d \pi_{od,t}^k \iota^k \left[\sum_k w_{d,t}^{k,G} L_{d,t}^{k,G} + \overline{\Pi_{d,t}^k} \right] \\ \implies \frac{1+\theta}{\theta} w_{o,t}^{k,G} L_{o,t}^{k,G} &= \sum_d \pi_{od,t}^k \iota^k \left[\sum_k \left(w_{d,t}^{k,G} L_{d,t}^{k,G} + \sum_r \varphi_{dr,t}^k \Pi_{r,t}^{k\;*} \right) \right] \\ \implies \frac{1+\theta}{\theta} w_{o,t}^{k,G} L_{o,t}^{k,G} &= \sum_d \pi_{od,t}^k \iota^k \left[\sum_k \left(w_{d,t}^{k,G} L_{d,t}^{k,G} + \sum_r \varphi_{dr,t}^k \frac{1}{\theta} w_{r,t}^{k,G} L_{r,t}^{k,G} \right) \right] \end{aligned} \tag{53}$$

where $\varphi_{dr,t}^k$ are the idea adoption shares and $\pi_{od,t}^k$ are the trade shares. Note that relative to the market clearing condition in Eaton and Kortum (2024), I eliminate the profit terms 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.6 Proposition 1 (Invariant Markup Distribution)

Proof. In each sector k, the probability that an idea i of quality q discovered at time t^* in region r undercuts the lowest cost competition in region d by a factor m at time t is:

$$\mathbb{P}\left[\frac{1}{q_{r,t^*}} \max_{o'} \left\{ \frac{w_{o',t}^k \tau_{o'd,t}^k}{a_{o',t}} \right\} \leq \frac{c}{m} \right] = \mathbb{P}\left[c \geq \frac{m}{q_{r,t^*}} \max_{o'} \left\{ \frac{w_{o',t}^k \tau_{o'd,t}^k}{a_{o',t}} \right\} \right] = 1 - G_d^k \left(\frac{m}{q_{r,t^*}} \max_{o'} \left\{ \frac{w_{o',t}^k \tau_{o'd,t}^k}{a_{o',t}} \right\} \right) \tag{54}$$

The conditional cost distribution in region d buying goods from region o is given by:

$$G_{od}^k(c) = \mathbb{P}\left[Z_o^k \geq \frac{w_o^k \tau_{od}^k}{c} \right] = 1 - F_o^k \left(\frac{w_o^k \tau_{od}^k}{c} \right) \tag{55}$$

Let $C_d(Z) = \min_{o'} C_{o'd}(Z)$. Then the unconditional cost distribution is:

$$\begin{aligned} G_d^k(c) &= \mathbb{P}\left[C_d(Z) \le c\right] = 1 - \mathbb{P}\left[C_d(Z) \ge c\right] \\ &= 1 - \exp\left\{ - \left[\sum_{r'=1}^{N} \int_{-\infty}^{t} \left[\sum_{o'=1}^{N} \Omega_{r'o',t^*}(t-t^*) \left(w_{o',t}^k \tau_{o'd,t}^k \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho} \lambda_{r',t^*}^k dt^* \right] c^\theta \right\} \\ &= 1 - \exp\left\{ - \left[\sum_{r'=1}^{N} \left[\sum_{o'=1}^{N} T_{r'o',t}^k \left(w_{o',t}^k \tau_{o'd,t}^k \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho} \right] c^\theta \right\} \end{aligned} \tag{56}$$

This is analogous to the unconditional cost distribution in Eaton and Kortum (2002) but with productivity drawn from a multivariate Fréchet distribution.

Given the unconditional cost distribution, I now build on Eaton and Kortum (2001) and compute the probability that an idea discovered in region r at time t^* will undercut the lowest cost competitor in region d at time t by m:

$$\begin{aligned} b_{rd}^k(m, t^*, t) &= \int_1^\infty \int_{\mathbb{R}^+} 1 - G_d^k\left(\frac{m}{q_{r,t^*}} \max_{o'} \left\{ \frac{w_{o',t}^k \tau_{o'd,t}^k}{a_{o',t}} \right\} \right) \; dM_r(a_1, \ldots, a_N; t-t^*) dH(q) \\ &= \int_1^\infty \int_{\mathbb{R}_+^N} \exp\left\{ - \left[\sum_{r'=1}^{N} \left[\sum_{o'=1}^{N} T_{r'o',t}^k \left(w_{o',t}^k \tau_{o'd,t}^k \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho} \right] \left(\frac{m}{q} \max_{o'} \left\{ \frac{w_{o',t}^k \tau_{o'd,t}^k}{a_{o'}} \right\} \right)^\theta \right\} \\ &\qquad dM_r(a_1, \ldots, a_N; t-t^*) dH(q) \\ &\approx \frac{m^{-\theta} \int_{\mathbb{R}_+^N} \max_{o'} a_{o'}^\theta \left(w_{o',t}^k \tau_{o'd,t}^k \right)^{-\theta} dM_r(a_1, \ldots, a_N; t-t^*)}{\sum_{r'=1}^N \left[\sum_{o'=1}^N T_{r'o',t}^k \left(w_{o',t}^k \tau_{o'd,t}^k \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho}} \\ &= \frac{m^{-\theta} \left[\sum_{o'=1}^N \Omega_{ro,t^*}(t-t^*) \left(w_{o',t}^k \tau_{o'd,t}^k \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho}}{\sum_{r'=1}^N \left[\sum_{o'=1}^N T_{r'o',t}^k \left(w_{o',t}^k \tau_{o'd,t}^k \right)^{-\frac{\theta}{1-\rho}} \right]^{1-\rho}} \\ &= \frac{m^{-\theta} \phi_{rd,t^*t}^k}{\lambda_{r,t^*}^k} \end{aligned} \tag{57}$$

where ϕ_{rd,t^*t} is defined in equation 17; the first equality applies the Marking Theorem for Poisson processes in Kingman (1992) – first used by Lind and Ramondo (2023a) to derive the multivariate Fréchet distribution for *productivity*; the approximation in the third line comes from how levels of q below 1 are handled, following equation (10) and Footnote 9 in Eaton and Kortum (2001), and; the equality in the fourth line comes from the joint distribution of applicabilities $M_r(a_1, \ldots, a_N; t-t^*)$ being independent Fréchet across o, as described in Appendix A2 in Lind and Ramondo (2024).

Hence the probability that an idea of quality q discovered in region r at time t^* will undercut the lowest cost competitor in region d at time t by m via production in region o is simply $b_{rd}(m, t^*, t)$ multiplied by the probability that region o is the least cost producer of the idea:

$$b_{rod}(m, t^*, t) = b_{rd}(m, t^*, t)\varphi_{rod|rd,t^*t} = \frac{m^{-\theta}\varphi_{rod|rd,t^*t}\phi_{rd,t^*t}}{\lambda_{r,t^*}} \tag{58}$$

Consequently, the probability that an idea discovered in region r at time t^* will lead to the respective good being sold in market d at t is:

$$b_{rd}(1, t^*, t) = \frac{\phi_{rd,t^*t}}{\lambda_{r,t^*}}. \tag{59}$$

Intuitively, ϕ_{rd,t^*t} is the share of goods produced in region t using ideas discovered in region r at time t^*, while λ_{r,t^*} is the arrival rate of ideas at time t^*. Thus, its ratio provides the probability that any given idea discovered in region r at time t^* will enter market d at t. Likewise, the probability that an idea discovered in region r at time t^* will enter market d at t via production in region o is:

$$b_{rod}(1, t^*, t) = \frac{\varphi_{rod|rd,t^*t}\phi_{rd,t^*t}}{\lambda_{r,t^*}}. \tag{60}$$

Thus, the markup distribution in region d at time t for a cohort of ideas from region r, conditional on selling in region d, **either conditional or unconditional on the production location** o, is time- and region-invariant and given by:

$$\mathbb{P}\left(M \leq m | M \geq 1\right) = \frac{b_{rd}(1, t^*, t) - b_{rd}(m, t^*, t)}{b_{rd}(1, t^*, t)} = \frac{b_{rod}(1, t^*, t) - b_{rod}(m, t^*, t)}{b_{rod}(1, t^*, t)} = 1 - m^{-\theta} = H(m). \tag{61}$$

□

C.7 Proposition 2 (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:

$$\frac{\omega^{k,R}_{r,t^*}}{\omega^{k',R}_{r,t^*}} = \underbrace{\frac{A^k_{t^*}}{A^{k'}_{t^*}}}_{\substack{\text{differences in}\\ \text{fundamental}\\ \text{research}\\ \text{productivity}\\ \text{across sectors}}} \cdot \underbrace{\frac{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\phi^k_{rd,t^*t}}{\lambda^k_{r,t^*}} \frac{X^k_{d,t}}{1+\theta} \frac{1}{P_{r,t}} dt}{\int_{t^*}^{\infty} e^{-\zeta(t-t^*)} \sum_{d=1}^{N} \frac{\phi^{k'}_{rd,t^*t}}{\lambda^{k'}_{r,t^*}} \frac{X^{k'}_{d,t}}{1+\theta} \frac{1}{P_{r,t}} dt}}_{\text{expected market potential of an idea}}. \tag{62}$$

The first term captures the compositional shift of innovation towards ICT at a national level, and the second term captures differences in the expected market potential of ideas across sectors.

C.8 Proposition 3 (Balanced Growth Path)

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

$$\lambda_r^k(t) = \gamma_r^k T_r^k(t) \implies \dot{\lambda}_r^k(t) = \gamma_r^k \dot{T}_r^k(t). \tag{63}$$

From equation 14 we know that the technology level at each time t is given by:

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

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

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

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

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

Thus we have that:

$$\begin{aligned}
\dot{T}_{o,t}^k &= \sum_{r=1}^{N} \gamma_r^k \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} \frac{\dot{T}_r^k(t^*)}{T_l^k(t^*)} T_r^k(t^*) dt^* \\
&= g \sum_{r=1}^{N} \gamma_r^k \int_{-\infty}^{t} \Omega_{ro}(t-t^*)^{1-\rho} T_r^k(t) e^{-g(t-t^*)} dt^* \\
&= \sum_{r} \gamma_r^k T_{r,t}^k \int_{-\infty}^{t} g e^{-g(t-t^*)} \left[1 - e^{-\delta_{ro}(t-t^*)} \right]^{1-\rho} dt^* \\
&= \sum_{r} \gamma_r^k T_{r,t}^k \int_{0}^{\infty} g e^{-ga} \left[1 - e^{-\delta_{ro}(a)} \right]^{1-\rho} da
\end{aligned} \tag{64}$$

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:

$$g\mathbf{T}^k = \mathbf{\Delta}^k(g)\mathbf{T}^k \tag{65}$$

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

$$\Delta_{ro}^k(g) = \gamma_r^k \int_0^\infty g e^{-ga} \left[1 - e^{-\delta_{ro}(a)}\right]^{1-\rho} da.$$

The aggregate growth rate is the Perron-Frobenius root of equation 65 with relative technology levels $\mathbf{T}$ corresponding to the Perron-Frobenius eigenvector that is defined up to a scalar multiple. Thus, any set of exogenous diffusion speeds δ_{ro} and endogenous innovation rates γ_r^k 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 in the general formulation with exponential diffusion and idea applicabilities. Suppose that the distribution of workers across regions, sectors, and occupations are constant on the BGP. Since the relative technology levels are also constant on the balanced growth path, the trade shares given by equation 17 and the market condition given by equation 34 imply that relative wages across regions and sectors in production is constant. In particular, trade shares on the BGP are given by:

$$\pi_{od}^k = \sum_{r=1}^N \int_0^\infty \frac{\Omega_{ro}(a) \left(w_o^k \tau_{od}^k\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^N \Omega_{ro'}(a) \left(w_{o'}^k \tau_{o'd}^k\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\gamma_r^k T_r^k e^{-g^k \cdot a} \left[\sum_{o'=1}^N \Omega_{ro'}(a) \left(w_{o'}^k \tau_{o'd}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^N \gamma_{r'}^k T_{r'}^k \int_0^\infty \left[\sum_{o'=1}^N \Omega_{r'o'}(a') \left(w_{o'}^k \tau_{o'd}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} e^{-g^k \cdot a'} da'} da. \tag{66}$$

where T_r^k are the relative technology levels determined by equation 65.

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

$$P_{d,t}^{k\;BGP} = \Gamma \left[\sum_{r'=1}^N \gamma_{r'}^k T_{r',t}^k \int_0^\infty e^{-ga} \left[\sum_{o'=1}^N \Omega_{r'o'}(a) \left(w_{o'}^k \tau_{o'd}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} da\right]^{-\frac{1}{\theta}} \tag{67}$$

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

$$g_{P^k} = \frac{\dot{P}_{d,t}^{k\;BGP}}{P_{d,t}^{k\;BGP}} = -\frac{1}{\theta} \sum_{r'} \breve{\pi}_{r'd} \frac{\dot{T}_{r',t}^k}{T_{r',t}^k} = -\frac{g^k}{\theta} \tag{68}$$

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

$$P_{d,t} = \check{\Gamma} \prod_k \left(P_{d,t}^k\right)^{\iota^k} \tag{69}$$

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

$$g_p = -\frac{1}{\theta} \sum_k \iota^k g^k \tag{70}$$

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

$$\begin{aligned}
\check{V}^k_{r,t^*} &= \frac{P_{r,t^*}}{1+\theta}\int e^{-\zeta(t-t^*)}\sum_d \frac{\phi_{rd,t^*t}}{\gamma^k_r T^k_{r,t^*}}\frac{X_{d,t}}{P_{r,t}}dt \\
&= \frac{\sum_d \phi_{rd}X_d}{1+\theta}\frac{1}{\gamma^k_r T^k_{r,t^*}}\int e^{-\zeta(t-t^*)}e^{-g_P(t-t^*)}dt \\
&= \frac{\sum_d \phi_{rd}X_d}{1+\theta}\frac{1}{\gamma^k_r T^k_{r,t^*}}\frac{1}{\zeta - g_P/\theta}.
\end{aligned} \tag{71}$$

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

$$w^{k,R}_r = \frac{\sum_d \check{\pi}_{rd}X_d}{1+\theta}\frac{1}{L^{k,R}_r}\frac{1}{\zeta - g_P/\theta}. \tag{72}$$

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

$$\begin{aligned}
V^{k,h}_{d,t} &= \widetilde{V}^{k,h}_{d,t} + g_V t \\
&= \log\left(\frac{w^{k,h}_{d,t}}{\widetilde{P}_{d,t}}\right) - g_P t + \frac{1}{\Upsilon}\log\left[\sum_o\sum_s\left(\sum_n \exp\left(\frac{1}{1+\zeta}\left[\widetilde{V}^{s,n}_{o,t'} + g_V t'\right] - \kappa^{ks,hn}_{do}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right] \\
&= \log\left(\frac{w^{k,h}_{d,t}}{\widetilde{P}_{d,t}}\right) - g_P t + \frac{1}{1+\zeta}g_V t' + \frac{1}{\Upsilon}\log\left[\sum_o\sum_s\left(\sum_n \exp\left(\frac{1}{1+\zeta}\widetilde{V}^{s,n}_{o,t'} - \kappa^{ks,hn}_{do}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right]
\end{aligned} \tag{73}$$

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:

$$\begin{aligned}
g_v &= -g_p + \frac{1}{1+\zeta}g_v \\
\implies g_v &= -\frac{1+\zeta}{\zeta}g_p = \frac{1+\zeta}{\zeta}\frac{1}{\theta}\sum_k \iota^k g^k
\end{aligned} \tag{74}$$

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 73 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:

$$\widetilde{V}^{k,h}_{d,t} = \log\left(\frac{w^{k,h}_{d,t}}{\widetilde{P}_{d,t}}\right) + \frac{1}{1+\zeta}g_v + \frac{1}{\Upsilon}\log\left[\sum_o\sum_s\left(\sum_n \exp\left(\frac{1}{1+\zeta}\widetilde{V}^{s,n}_{o,t'} - \kappa^{ks,hn}_{do}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right]. \tag{75}$$

Substituting the decomposition of expected worker value in equation 73 into equation 32 in the main text, worker mobility shares are alternately given by:

$$\mu_{do,t}^{ks,hn} = \frac{\exp\left(\frac{1}{1+\zeta}\widetilde{V}_{o,t'}^{s,n} - \kappa_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum\limits_{n'}\exp\left(\frac{1}{1+\zeta}\widetilde{V}_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}} \cdot \frac{\left[\sum\limits_{n'}\exp\left(\frac{1}{1+\zeta}\widetilde{V}_{o,t'}^{s,n'} - \kappa_{do,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum\limits_{o'}\sum\limits_{s'}\left[\sum\limits_{n'}\exp\left(\frac{1}{1+\zeta}\widetilde{V}_{o',t'}^{s',n'} - \kappa_{do',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}} \tag{76}$$

and are constant on the BGP. Hence, from equation 33, 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 65, 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 $g_v = \frac{1+\zeta}{\zeta}\frac{1}{\theta}\sum_k \iota^k g^k$. $\square$

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 (64) 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:

$$\pi_{od,t}^{NA} = \frac{T_{o,t}\left(w_{o,t}\tau_{od,t}\right)^{-\theta}}{\sum\limits_{o'=1}^{N} T_{o',t}\left(w_{o',t}\tau_{o'd,t}\right)^{-\theta}}, \quad P_{d,t}^{NA} = \Gamma\left[\sum_{o'=1}^{N} T_{o',t}\left(w_{o',t}\tau_{o'd,t}\right)^{-\theta}\right]^{-\frac{1}{\theta}} \tag{77}$$

and idea diffusion shares are given by:

$$\varphi_{ro,t}^{NA} = \sum_{d=1}^{N}\frac{T_{ro,t}\left(w_{o,t}\tau_{od,t}\right)^{-\theta}}{\sum\limits_{o'=1}^{N} T_{o',t}^{k}\left(w_{o',t}\tau_{o'd,t}\right)^{-\theta}}. \tag{78}$$

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 (64) collapses to:

$$\begin{aligned}\dot{T}^k_{o,t} &= \sum_r \gamma^k_r T^k_{r,t} \int_0^\infty g e^{-ga} \delta_{ro}^{1-\rho} da \\ &= \sum_r \delta_{ro}^{1-\rho} \gamma^k_r T^k_{r,t}.\end{aligned}$$

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

$$\pi^{ID}_{od,t} = \sum_{r=1}^N \frac{(T_{ro,t})^{\frac{1}{1-\rho}} (w_{o,t}\tau_{od,t})^{-\frac{\theta}{1-\rho}}}{\sum\limits_{o'=1}^N \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum\limits_{o'=1}^N \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum\limits_{r'=1}^N \left[\sum\limits_{o'=1}^N \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{79}$$

$$P^{ID}_{d,t} = \Gamma \left[\sum_{r'=1}^N \left[\sum_{o'=1}^N \left(T_{r'o',t}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\right]^{-\frac{1}{\theta}} \tag{80}$$

and the idea diffusion shares are:

$$\varphi^{ID}_{od,t} = \frac{(T_{ro,t})^{\frac{1}{1-\rho}} (w_{o,t}\tau_{od,t})^{-\frac{\theta}{1-\rho}}}{\sum\limits_{o'=1}^N \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum\limits_{o'=1}^N \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum\limits_{l'=1}^N \left[\sum\limits_{o'=1}^N \left(T_{ro',t}\right)^{\frac{1}{1-\rho}} \left(w_{o',t}\tau_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{81}$$

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

C.9 Proposition 4 (Transition Path)

Proof. The transition path towards balanced growth is characterized by a trajectory of detrended expected values given by equation 75, worker mobility shares given by equation 76 and evolution of worker population given by equation 33:

$$\begin{aligned}\widetilde{V}^{k,h}_{d,t} &= \log\left(\frac{w^{k,h}_{d,t}}{\widetilde{P}_{d,t}}\right) + \frac{1}{1+\zeta} g_v + \frac{1}{\Upsilon} \log\left[\sum_o \sum_s \left(\sum_n \exp\left(\frac{1}{1+\zeta}\widetilde{V}^{s,n}_{o,t'} - \kappa^{ks,hn}_{do}\right)^{\frac{\Upsilon}{\upsilon}}\right)^{\upsilon}\right] \\ \mu^{ks,hn}_{do,t} &= \frac{\exp\left(\frac{1}{1+\zeta}\widetilde{V}^{s,n}_{o,t'} - \kappa^{ks,hn}_{do,t}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum\limits_{n'} \exp\left(\frac{1}{1+\zeta}\widetilde{V}^{s,n'}_{o,t'} - \kappa^{ks,hn'}_{do,t}\right)^{\frac{\Upsilon}{\upsilon}}} \cdot \frac{\left[\sum\limits_{n'} \exp\left(\frac{1}{1+\zeta}\widetilde{V}^{s,n'}_{o,t'} - \kappa^{ks,hn'}_{do,t}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum\limits_{o'}\sum\limits_{s'} \left[\sum\limits_{n'} \exp\left(\frac{1}{1+\zeta}\widetilde{V}^{s',n'}_{o',t'} - \kappa^{ks',hn'}_{do',t}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}} \\ L^{s,n}_{o,t'} &= \sum_h \sum_k \sum_d \mu^{ks,hn}_{do,t} L^{k,h}_{d,t},\end{aligned}$$

where at each time t the innovation and technology levels are given by equations 8 and 14 respectively:

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

the trade shares, idea adoption shares, price index, and market clearing condition are given by equations 17, 22, 21, and 34 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 26, the returns to innovation by equation 25 and total expenditures by 47 and idea market shares by 48:

$$w_{r,t}^{k,R} = A_{t}^{k} A_{r,t} \left(L_{r,t}^{k,R}\right)^{\alpha} T_{r,t}^{k} \check{V}_{r,t}^{k} = \frac{\check{V}_{r,t}^{k} \lambda_{r,t}^{k}}{L_{r,t}^{k,R}}$$

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

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

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

The detrended equations for the migration part of the model (i.e. equations 75 and 76) can be further expressed using dynamic hat algebra (Caliendo et al., 2019). Let $\widehat{x}_{t'} = \frac{\widetilde{x}_{t'}}{\widetilde{x}_t}$ for any variable x and $u = \exp(V)$. Then we have:

$$\log\left(\widehat{u}_{d,t}^{k,h}\right) = \log\left(\frac{\widehat{w}_{d,t}^{k,h}}{\widehat{P}_{d,t}}\right) + \frac{1}{\Upsilon}\log\left[\sum_n\left[\sum_s\sum_o \mu_{do,t'}^{ks,hn}\left(\widehat{u}_{o,t'}^{s,n}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{do,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}\right] \tag{82}$$

$$\mu_{od,t'}^{ks,hn} = \frac{\mu_{od,t}^{ks,hn}\left(\widehat{u}_{d,t'}^{s,n}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn}\right)^{\frac{\Upsilon}{\upsilon}}}{\sum_{n'}\mu_{od,t}^{ks,hn'}\left(\widehat{u}_{d,t'}^{s,n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}} \cdot \frac{\left[\sum_{n'}\mu_{od,t}^{ks,hn'}\left(\widehat{u}_{d,t'}^{s,n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od,t}^{ks,hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}}{\sum_{n'}\left[\sum_{d'}\sum_{s'}\mu_{od',t}^{ks',hn'}\left(\widehat{u}_{d',t'}^{s',n'}\right)^{\frac{\Upsilon}{(1+\zeta)\upsilon}}\left(\widehat{\kappa}_{od',t}^{ks',hn'}\right)^{\frac{\Upsilon}{\upsilon}}\right]^{\upsilon}} \tag{83}$$

$$L_{d,t'}^{k,h} = \sum_n\sum_s\sum_o \mu_{od,t}^{ks,hn} L_{o,t}^{s,n}. \tag{84}$$

Data on trade shares and innovation levels are at an annual frequency. To provide a direct mapping between my model and the data, I make additional assumptions that govern the timing of innovation and production. While inventors receive ideas from a Poisson process described in the main text, a separate Poisson arrival process with rate 1 governs when each inventor can produce those ideas. Another independent Poisson arrival process with rate 1 governs when each firm in each region can produce goods.

Given these assumptions, technology levels can be expressed as:

$$T_{o,t}^k = \sum_{r=1}^N\sum_{t^*\in\mathfrak{T}_t}\Omega_{ro,t^*}(t-t^*)^{1-\rho}\cdot\lambda_{r,t^*}^k + T_{o,0}^k \tag{85}$$

where $\mathfrak{T}_t$ are the set of innovation times t^* from time 0 to time t. Though these innovation times are stochastic, the expected interval between any two innovation times is 1. Note that the diffusion term $\Omega_{ro,t^*}(t-t^*)$ remains unchanged. The innovation levels remain unchanged from equation 8. The trade shares, idea diffusion shares, and price index can be expressed as:

$$\pi_{od,t}^k = \sum_{r=1}^N\sum_{t^*\in\mathfrak{T}}\frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}^k\tau_{od,t}^k\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^k\tau_{o'd,t}^k\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^k\tau_{o'd,t}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^*}^k}{\sum_{r'}\sum_{t^*\in\mathfrak{T}}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^k\tau_{o'd,t}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^k} \tag{86}$$

$$\varphi_{ro,t}^k = \sum_{d=1}^N\sum_{t^*\in\mathfrak{T}}\frac{\Omega_{ro,t^*}(t-t^*)\left(w_{o,t}^k\tau_{od,t}^k\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^k\tau_{o'd,t}^k\right)^{-\frac{\theta}{1-\rho}}}\frac{\left[\sum_{o'}\Omega_{ro',t^*}(t-t^*)\left(w_{o',t}^k\tau_{o'd,t}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r,t^*}^k}{\sum_{r'}\sum_{t^*\in\mathfrak{T}}\left[\sum_{o'}\Omega_{r'o',\check{t}}(t-\check{t})\left(w_{o',t}^k\tau_{o'd,t}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',\check{t}}^k} \tag{87}$$

$$P_{d,t}^k = \Gamma\left[\sum_{r'=1}^N\sum_{t^*\in\mathfrak{T}}\left[\sum_{o'=1}^N\Omega_{r'o',t^*}(t-t^*)\left(w_{o',t}^k\tau_{o'd,t}^k\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\lambda_{r',t^*}^k\right]^{-\frac{1}{\theta}}, \tag{88}$$

while the market clearing condition remains unchanged from equation 34. Inventor wages are given by equation 26, total expenditures by equation 47, while the value of an idea and the idea market shares are now given by:

$$\check{V}^k_{r,t} = \sum_{t' \in \mathcal{T}^{\infty}_t} \left(\frac{1}{1+\zeta}\right)^{t'-t} \sum_{d=1}^{N} \frac{X^k_{d,t'}}{1+\theta} \cdot \frac{P_{r,t}}{P_{r,t'}} \cdot \frac{\phi^k_{rd,tt'}}{\lambda^k_{r,t}} \tag{89}$$

$$\phi^k_{rd,tt'} = \frac{\left[\sum_{o'} \Omega_{ro',t}(t'-t) \left(w^k_{r',t'} \tau^k_{ro',t'}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} \lambda^k_{r,t}}{\sum_{r'} \sum_{\check{t} \in \mathcal{T}^{t'}_{-\infty}} \left[\sum_{o'} \Omega_{r'o',\check{t}}(t'-\check{t}) \left(w^k_{o',t'} \tau^k_{o'd,t'}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho} \lambda^k_{r',\check{t}}}. \tag{90}$$

$\square$

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 77 and 78:

$$\pi^{k,NA}_{od,t} = \frac{T^k_{o,t}\left(w^k_{o,t}\tau^k_{od,t}\right)^{-\theta}}{\sum_{o'=1}^{N} T^k_{o',t}\left(w^k_{o',t}\tau^k_{o'd,t}\right)^{-\theta}}, \quad \varphi^{k,NA}_{ro,t} = \sum_{d=1}^{N} \frac{T^k_{ro,t}\left(w^k_{o,t}\tau^k_{od,t}\right)^{-\theta}}{\sum_{o'=1}^{N} T^k_{o',t}\left(w^k_{o',t}\tau^k_{o'd,t}\right)^{-\theta}}, \quad P^{k,NA}_{d,t} = \Gamma\left[\sum_{o'=1}^{N} T^k_{o',t}\left(w^k_{o',t}\tau^k_{o'd,t}\right)^{-\theta}\right]^{-\frac{1}{\theta}}.$$

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:

$$\widehat{\pi}^{k,NA}_{od,t} = \frac{\widehat{T}^k_{o,t}\left(\widehat{w}^k_{o,t}\widehat{\tau}^k_{od,t}\right)^{-\theta}}{\sum_{o'=1}^{N} \widehat{T}^k_{o',t}\left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\theta}} \tag{91}$$

$$\widehat{\varphi}^{k,NA}_{ro,t} = \sum_{d=1}^{N} \frac{\widehat{T}^k_{ro,t}\left(\widehat{w}^k_{o,t}\widehat{\tau}^k_{od,t}\right)^{-\theta}}{\sum_{o'=1}^{N} \widehat{T}^k_{o',t}\left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\theta}} \tag{92}$$

$$\widehat{P}^{k,NA}_{d,t} = \left[\sum_{o'=1}^{N} \widehat{T}^k_{o',t}\left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\theta}\right]^{-\frac{1}{\theta}} \tag{93}$$

and the market clearing condition remains unchanged from equation 34. Innovation levels remain unchanged from equation 8 and changes in technology levels are given by:

$$\begin{aligned} T^k_{o,t'} - T^k_{o,t} &= \sum_{r} \delta_{ro,t}\left(\Lambda^k_{l,t} - T^k_{lo,t}\right) \\ \implies \widehat{T}^k_{o,t'} &= 1 - \delta_{ro,t} + \sum_{l} \frac{\delta_{lo,t}\Lambda^k_{l,t}}{T^k_{o,t}} \end{aligned} \tag{94}$$

where $\Lambda^k_{r,t} = \sum_{t^* \in \mathcal{T}} \lambda^k_{r,t^*}$ 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 79 -81:

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

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

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

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

$$\widehat{\pi}^{k,ID}_{od,t} = \sum_{r=1}^{N} \frac{\left(\widehat{T}^k_{ro,t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o,t}\widehat{\tau}^k_{od,t}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N} \left(\widehat{T}^k_{ro',t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N} \left(\widehat{T}^k_{ro',t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N} \left[\sum_{o'=1}^{N} \left(\widehat{T}^k_{ro',t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{95}$$

$$\widehat{P}^{k,ID}_{d,t} = \left[\sum_{r'=1}^{N} \left[\sum_{o'=1}^{N} \left(\widehat{T}^k_{r'o',t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}\right]^{-\frac{1}{\theta}} \tag{96}$$

$$\widehat{\varphi}^{k,ID}_{od,t} = \frac{\left(\widehat{T}^k_{ro,t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o,t}\widehat{\tau}^k_{od,t}\right)^{-\frac{\theta}{1-\rho}}}{\sum_{o'=1}^{N} \left(\widehat{T}^k_{ro',t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}} \cdot \frac{\left[\sum_{o'=1}^{N} \left(\widehat{T}^k_{ro',t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}}{\sum_{r'=1}^{N} \left[\sum_{o'=1}^{N} \left(\widehat{T}^k_{ro',t}\right)^{\frac{1}{1-\rho}} \left(\widehat{w}^k_{o',t}\widehat{\tau}^k_{o'd,t}\right)^{-\frac{\theta}{1-\rho}}\right]^{1-\rho}} \tag{97}$$

and the market clearing condition remains unchanged from equation 34. Changes in technology levels are given by:

$$\begin{aligned} T^k_{o,t'} - T^k_{o,t} &= \sum_r \delta^{1-\rho}_{ro,t} \gamma^k_{r,t} T^k_{r,t} \\ \implies \widehat{T}^k_{o,t'} &= 1 + \sum_r \delta^{1-\rho}_{ro,t} \gamma^k_{r,t} \frac{T^k_{r,t}}{T^k_{o,t}} \end{aligned} \tag{98}$$

and innovation levels remain unchanged from equation 8.

C.10 Corollary 3 (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 (31) 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-migration 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:

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

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:

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

where $\ddot{x}_{t'} = \frac{\widehat{\acute{x}_{t'}}}{\widehat{x}_t}$ denotes the counterfactual change in any detrended variable x. Recall that $\widehat{x}_{t'} = \frac{\widetilde{x}_{t'}}{\widetilde{x}_t}$ denotes the time changes in any detrended variable x.

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

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

□

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, equations 28-33 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:

$$L_{o,t}^{k} = \left[\left(L_{o,t}^{H,k} \right)^{\frac{\varphi-1}{\varphi}} + \left(L_{o,t}^{S,k} \right)^{\frac{\varphi-1}{\varphi}} \right]^{\frac{\varphi}{\varphi-1}} \tag{99}$$

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:

$$Y_{o,t}^{k}(\nu) = z_{o,t}^{k}(\nu) \left[(K_{o,t}^{k})^{\psi} (L_{o,t}^{k})^{1-\psi} \right]^{\chi} M_{o,t}^{1-\chi} \tag{100}$$

where $z_{o,t}(\nu)$ is the productivity drawn from the multivariate Fréchet distribution given by equation 12, $K_{o,t}$ 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:

$$x_{o,t}^{k} = \check{\Gamma} \left[(\check{r}_{o,t})^{\psi} \left(w_{o,t}^{k} \right)^{1-\psi} \right]^{\chi} P_{o,t}^{k\;1-\chi} \tag{101}$$

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 17 yields the equilibrium trade shares.

Capital market clearing is given by:

$$\check{r}_{o,t} K_{o,t}^{k} = \frac{1-\psi}{\psi} w_{o,t}^{k} L_{o,t}^{k} \tag{102}$$

Since capital income is a constant multiple of production worker income, the combined capital and labor market clearing condition is still given by equation 34 in the main paper. With input-output loops of the form in equation 101 – where goods producers purchase the final good only in that sector[25] – intermediate good spending is a constant multiple of production worker and capital income:

$$X_{o,t}^{I} = \frac{1-\chi}{\chi} \left(w_{o,t}^{k} L_{o,t}^{k} + r_{o,t} K_{o,t}^{k} \right) = \frac{1-\chi}{\chi} \frac{1}{\psi} w_{o,t}^{k} L_{o,t}^{k}. \tag{103}$$

Thus, the combined capital and labor market clearing condition is still given by equation 34 in the main

[25]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:

$$\check{V}_{o,t}^K = \mathbb{E}_t \sum_{s=0}^{\infty} \beta^{t+s} \frac{\left(C_{o,t+s}^K\right)^{1-1/\eta}}{1-1/\eta} \tag{104}$$

subject to their budget constraint:

$$r_{o,t} K_{o,t} = P_{o,t} \left(C_{o,t}^K + K_{o,t+1} - (1-\delta_{o,t}) K_{o,t}\right) \tag{105}$$

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, given by equation (28).

The optimal consumption and saving decisions are given by:

$$C_{o,t}^K = \varsigma_{o,t} R_{o,t} K_{o,t} \tag{106}$$

$$K_{o,t+1} = \left(1 - \varsigma_{o,t}\right) R_{o,t} K_{o,t} \tag{107}$$

$$\varsigma_{o,t}^{-1} = 1 + \beta^{\eta} \left(\mathbb{E}_t \left[R_{o,t+1}^{\frac{\eta-1}{\eta}} \varsigma_{o,t+1}^{-\frac{1}{\eta}}\right]\right)^{\eta} \tag{108}$$

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:

$$U\left(C_{o,t}, B_{o,t}\right) = \log\left(B_{o,t} C_{o,t}\right) \tag{109}$$

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 100 now becomes:

$$Y_{o,t}^k(\nu) = z_{o,t}^k(\nu) (L_{o,t}^k)^{\breve{\alpha}} \left[(K_{o,t}^k)^{\psi} (L_{o,t}^{k,G})^{1-\psi}\right]^{\chi} M_{o,t}^{1-\chi} \tag{110}$$

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