Right Idea, Wrong Place? Knowledge Diffusion and Spatial Misallocation in R&D*

Trevor C. Williams

Department of Economics, Yale University†

Job market paper (click for latest version)

November 1, 2022

Abstract

A few cities perform a large share of research and development (R&D) in the United States. To the extent that R&D generates local knowledge spillovers, then the social returns to R&D vary across cities and the geographic distribution of researchers may be inefficient. Equally important is how the private returns to R&D vary across cities, on which there is less evidence. To unpack variation in private returns, I document a new fact from the market for technology: patent sales from inventor to firm decline steeply with distance. Through the lens of a spatial growth model, I infer that the private returns to R&D are low in remote regions because it is hard to commercialize inventions in distant markets. By contrast, the social returns are flatter across space because patent citations depend less on distance. Place-based R&D policy subsidizes research in locations where private returns are low relative to social returns. The optimal policy increases patenting by 2.8% and aggregate consumption by 0.8% in the long run, with minimal effects on inequality across regions or workers.

---

*Preliminary and incomplete. Thanks to my advisors Costas Arkolakis, Giuseppe Moscarini, and Michael Peters. Thanks also to Treb Allen, Lorenzo Caliendo, Lucas Conwell, Cecilia Fieler, John Finlay, Teresa Fort, Sam Kortum, Song Ma, Ryungha Oh, Piyush Panigrahi, Ahyan Panjwani, Bernardo Ribeiro, Michael Sullivan, Lindsey Uniat, Kaushik Vasudevan, Conor Walsh, and seminar participants at Yale and the 2022 North American Meeting of the Urban Economics Association for their comments and suggestions.

†`trevor.c.williams@yale.edu`

1 Introduction

A few cities in the US patent much more than others. In 2019, Boston received three times as many patents per capita as Denver, a city with comparable size, average wage, and educational attainment. Is the geographic concentration of research and development (R&D) in cities like Boston efficient? The question hinges on how the social returns and private returns to R&D vary across cities. Prior literature in economics has emphasized local knowledge spillovers in R&D, suggesting the social returns are high in research hubs like Boston. There is less work on private returns. Inventors in Boston may enjoy higher private returns than inventors in Denver because Boston is close to customers throughout the Northeast, for example. On balance, the extent and direction of inefficiency is not clear. This paper examines place-based R&D policy: how to reallocate researchers to align the social and private returns to R&D, city by city. Should policymakers favor R&D in Denver versus Boston, or Boston versus Denver? What are the aggregate welfare gains, and how are the gains distributed across cities and workers?

Variation in the private and social returns to R&D arise from imperfect knowledge diffusion across space. I consider two distinct barriers to knowledge diffusion. Learning barriers, from inventor to inventor, govern the social returns. Selling barriers, from inventor to firm, shape private returns. Learning barriers have been documented in literature on patenting going back to Jaffe, Trajtenberg, and Henderson (1993), but selling barriers have been less widely studied, at least in the US context. I make three contributions. Empirically, I offer new evidence from patents data consistent with appreciable selling barriers within the US. Theoretically, I present a unified treatment of diffusion barriers in a model of spatial growth and characterize optimal R&D employment subsidies. Quantitatively, I implement the optimal budget-neutral tax and subsidy program. Policy can increase long-run aggregate consumption by just under 1%.

On the empirical side, I first establish that knowledge diffusion is imperfect. Patenting and average wages at the city level both increase with the stock of patents in nearby cities. A shift-share instrumental variable design, in which I interact lagged patenting in foreign countries with each city's historical foreign population ancestry, provides exogenous variation in patent stocks. A one standard deviation increase in neighbors' patent stocks increases local patenting by 0.15 standard deviations and local wages by 0.2 standard deviations.

Patterns of patent citations and patent sales support the hypothesis that inventors and firms have imperfect access to ideas created at a distance. Prior literature has shown, and I confirm, that patent citations between cities decline with distance. I add a new fact: patent sales decline steeply with distance, even conditional on cities' sectors and technology mixes. Sales fall off more steeply than citations. All else equal, an inventor in Denver is only 20% as likely to sell a patent to a firm in faraway Boston compared to a firm in nearby Colorado Springs. By comparison, the same inventor is 55% as likely to receive a patent citation from an inventor in Boston as from an inventor in Colorado Springs.

Next, I study how imperfect knowledge diffusion affects the private and social returns to research in a model of semi-endogenous growth. Researcher create new product varieties and collect profits when varieties are adopted by firms. Product designs are non-rival and can be adopted by firms in many regions. Adoption is limited by exogenous selling barriers. In parallel, research generates spillovers as in Romer (1990) and Jones (1995), whereby researchers draw on their local knowledge pool to build upon previous discoveries. An exogenous learning barrier reduces the probability that a given insight enters the local pool. As in the data, both diffusion barriers increase with distance at potentially different rates.

The private returns to research are high when outward selling barriers are low or when the home market is large. Denver's remoteness helps rationalize why it innovates relatively little despite its high share of college-educated workers. The social returns are high when outward learning barriers are low or when the number of local researchers is large. My model unifies elements of Eaton and Kortum (1999), which features selling barriers but perfect spillovers, and Grossman and Helpman (2018), which emphasizes learning barriers without variation in market size. Relative to those papers, my model focuses on the decisions of mobile researchers and I derive results for optimal policy. The structure of my model builds off of Walsh (2019), who formulates a tractable theory of local growth. Whereas I abstract from dynamics off the balanced growth path, which are the focus of Walsh (2019), I add regional interactions through a knowledge diffusion network.

The model delivers a sufficient statistic for the optimal research subsidy: the ratio of social returns, which encode knowledge spillovers, to private returns, which capture market access. Boston inspires large spillovers, hence high social returns, but it also collects large profits, hence high private returns. Denver, being remote from both inventors and other customers, has lower returns of both types.

The fact that sales decline more steeply with distance than citations suggests that variation in private returns dominates variation in social returns. This intuition is borne out by a quantitative version of the model calibrated to match the reduced-form analysis. Inventors in Denver struggle to monetize their ideas, and so there are too few inventors and too few ideas in there. But low private returns are not matched by low social returns, because researchers in Denver can still make adequate contributions to the knowledge pools in other regions. The intuition is easiest to see in the special case in which selling is purely local and learning is completely global. Then, a new variety created in a tiny region collects few profits, but is just as socially useful as a variety created in a big city. The optimal policy reallocates researchers to cities where the cost of producing an additional variety is low, which are places that tend to look like Denver.

I then relate my findings to real-world policy. Policymakers in the US are concerned about regional inequality in innovation. The CHIPS and Science Act of 2022 earmarked $10 billion for "regional technology hubs," cities which are not currently research-intensive but which in principle could develop an R&D ecosystem. I implement the CHIPS Act in my model, subsidizing research in cities with low patent output but high college shares. This policy is inefficiently blunt and aggregate consumption and wages fall slightly. Surprisingly, production wages fall even in the targeted hubs, because local increases in innovation are more than offset by lower spillovers from other cities. The example highlights the value of my quantitative model over a heuristic policy rule.

Prior literature

Existing work on knowledge diffusion focuses on the international context (Eaton and Kortum 1999; Keller 2002; Peri 2005; Keller and Yeaple 2013; Grossman and Helpman 2018), while I instead consider innovation across US regions. Desmet, Nagy, and Rossi-Hansberg (2018) build a model of spatial growth in which the returns to innovation depend on local activity, as in my paper. Relative to that work, I characterize how incentives vary with diffusion barriers and study optimal policy.

Papers in macro-development have attributed slow technology diffusion across countries to adoption costs (Comin and Hobijn 2010; Manuelli and Seshadri 2014), institutional barriers (Parente and Prescott 1994), or inappropriate technologies (Basu and Weil 1998; Acemoglu and Zilibotti 2001; Acemoglu 2002; Caselli and Coleman II 2006; Souza 2022; Moscona and Sastry 2022). Comin, Dmitriev, and Rossi-Hansberg (2012) ascribe slow diffusion to infrequent social interactions over long distances. Patent citations decline with distance, even within the same technology class (Jaffe, Trajtenberg, and Henderson 1993; Figueiredo, Guimarães, and Woodward 2015; Buzard et al. 2020; Berkes, Gaetani, and Mestieri 2021; Kwon et al. 2022) and inventors benefit from locating in dense research clusters (Kerr and Kominers 2015; Lychagin et al. 2016; Moretti 2021; Atkin, Chen, and

Popov 2022). Knowledge is embodied within people and travels with them as they move through space (Breschi and Lissoni 2009; Arkolakis, Lee, and Peters 2020; Cai et al. 2022; Prato 2022). There is less work on the spatial diffusion of knowledge from inventors to producers. Consistent with my findings, Hausman (2022) shows that wages, employment, and patenting increase at firms near universities following a policy-driven shock to university innovation. Bloom et al. (2021) document slow diffusion of new technologies in online job ads. The diffusion lags implied by my model, calibrated to patent sales data, are broadly in line with their estimates. Studies of patent sales have considered incentives on both sides of the market and documented characteristics of buyers and sellers (Arora and Fosfuri 2003; Serrano 2010; Akcigit, Celik, and Greenwood 2016; Figueroa and Serrano 2019). In complementary work, Arque-Castells and Spulber (2022) study how the market for patents shapes the private and social returns to R&D. Arora et al. (2022) show that that the private returns to R&D vary across firm size because larger firms can better commercialize their inventions, which is consistent with my proposed mechanism at a micro level. My paper focuses on spatial knowledge diffusion and policy.

Finally, my paper relates to the literature on R&D misallocation and policy (Lentz and Mortensen 2016; Acemoglu et al. 2018; Atkeson and Burstein 2019; Atkeson, Burstein, and Chatzikonstantinou 2019; Peters 2020; Akcigit, Hanley, and Serrano-Velarde 2021; König et al. 2022). These papers deal with misallocation either between research and production or across firms. An exception is Liu and Ma (2021), who study knowledge spillovers across sectors and on whose conceptual framework I build. Atkeson and Burstein (2019) argue that addressing misallocation in R&D is challenging because it is not obvious how to target the right activities or firms. One advantage of place-based R&D policy is that geography, while too general to capture all relevant features of innovation, is comparatively straightforward to measure. Moretti and Wilson (2014) study the effects of state-level R&D subsidies and Schweiger, Stepanov, and Zacchia (2022) examine placed-based R&D policy in the creation of Russian "Science Cities." Berkes and Gaetani (2021) analyze the socially optimal provision of "unconventional" innovation, in which dense cities have a comparative advantage. Policy-oriented papers (Gruber and Johnson 2019; Atkinson, Muro, and Whiton 2019; Glaeser and Hausman 2020) advocate subsidies to innovation outside major research hubs. I contribute to this literature by formalizing a new motive for efficient place-based R&D policy: imperfect knowledge diffusion.

The structure of the paper is as follows. Section 2 provides evidence of imperfect knowledge diffusion. Section 3 presents the model, followed by analysis of policy in Section 4. Section 5 enriches the model for the quantitative analysis of Section 6. Section 7 concludes.

2 Facts

I present three facts on the geography of patenting in the United States. One, patenting is concentrated: the number of patents per worker varies widely across cities. Two, the productivity effects of patents are local: city-level patenting and wages increase with the stock of patents in nearby cities. Three, the market for patents is local: inventors tend to sell patents to nearby firms.

The principal data sources are patents data from the United States Patent and Trademark Office (USPTO 2019) and labor market data from IPUMS (Ruggles et al. 2022). I work at the level of core-based statistical areas (CBSAs). The Census Bureau designates CBSAs as groups of counties sharing an urban core and connected by commuting ties, and I will refer to CBSAs as "cities" or "regions." In 2013 there were 917 CBSAs in the US, covering just under 90% of the US population.

I geocode each patent granted in the US between 1976 and 2019 to the inventors' city of residence. If a patent has multiple inventors living in different cities, I split the patent between cities giving equal weight to each inventor. Foreign inventors and inventors living outside of CBSAs are omitted from the analysis. I determine location based on inventors' residence rather than the firm's address to best capture where innovation occurs.[1]

Fact 1. *The number of patents granted per worker varies widely across cities.*

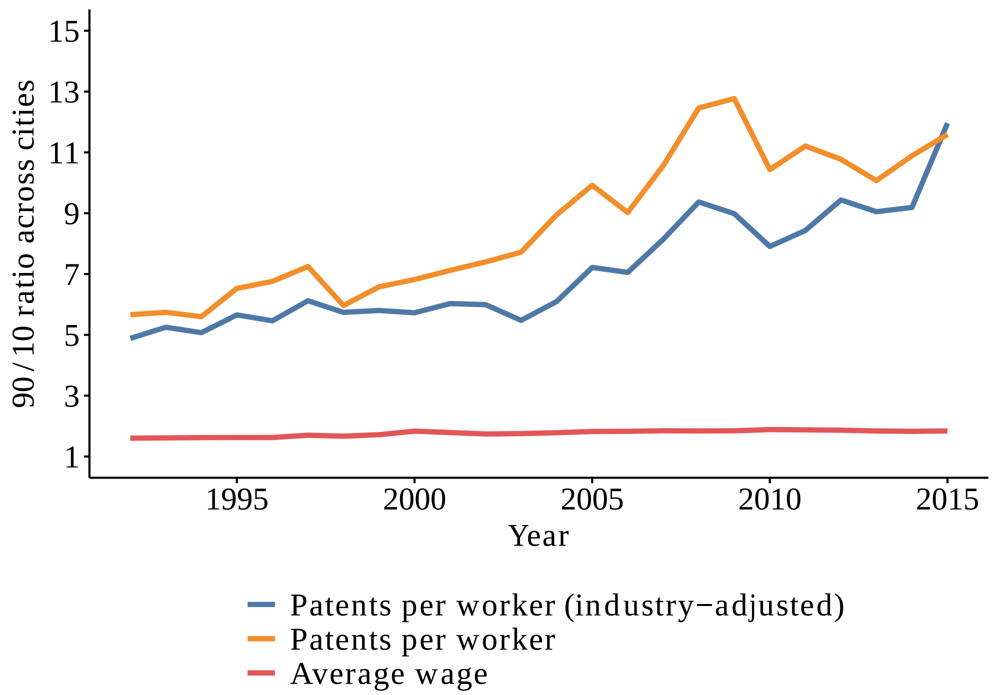
Patenting and R&D are geographically concentrated (Forman, Goldfarb, and Greenstein 2014; Andrews and Whalley 2022; Buzard et al. 2017). To summarize concentration, I compute the number of patents per worker granted in each city and year from 1990 to 2017. Figure 1 plots the 90/10 ratio of this distribution, where each percentile is weighted by employment.[2] Patenting is concentrated and has become more so over time. On average, the 90th percentile city creates over ten times as many patents per worker as the 10th percentile city. This advantage has doubled since the early 1990s. The 90/10 ratio of average wages, by comparison, rose from 1.6 to 1.85.

Patenting intensity varies across industries, so it is possible that the geographic distribution of patents reflects the geographic distribution of industries. To adjust for local industry

[1] Firm addresses in patent documents are often headquarters establishments. However, the number of patents by inventor city and year and the number of patents by firm city and year have a correlation of 0.94.

[2] I compute the 90/10 ratio on five-year moving averages.

Figure 1: Variation across cities in patents per worker and average wages



Patents per worker are five-year rolling averages. "Industry-adjusted" divides patents per worker by predicted patents per worker. Predicted patents per worker interacts the city's industry employment shares with national average patents per worker by industry.

composition, I compute predicted patents per worker with the formula

$$\widehat{\frac{\text{patents}}{\text{worker}}}_{j,t} = \sum_{ind} \frac{\text{national patents}_{ind}}{\text{national employment}_{ind}} \times \text{employment share}_{j,t,ind} \tag{1}$$

where j indexes cities, t indexes years, and *ind* indexes four-digit NAICS industries. The industry-adjusted measure in Figure 1 is patents per worker divided by predicted patents per worker. If the patent distribution were driven only by industry composition, then industry-adjusted patents per worker would be one in every location. Figure 1 shows this is not the case: dispersion in patents per worker is just as high when adjusting for industry mix. Appendix Figure A1 shows that the relative levels and trends are similar for the standard deviations of log wages and log patents per worker.

While I focus on patenting in this paper, the geographic concentration of innovation extends to R&D inputs. I measure business R&D spending from the National Science Foundation Business and Enterprise Research and Development Survey (National Science Foundation 2018) and venture capital from Crunchbase (2021). Appendix Figure A2 shows that 60% of

Table 1: Effect of neighbors' patent stock on city-level patents and wages

	Δ Log patents$_t$		Δ Log wage$_t$	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
Δ Neighbors' log patents$_{[t-10,t-1]}$	1.25^{***}	1.86^{***}	0.153^{***}	0.306^{***}
	(0.442)	(0.593)	(0.033)	(0.055)
Year FE	✓	✓	✓	✓
Observations	1,834	1,834	1,834	1,834
R^2	0.06	0.06	0.02	0.01
KP F stat		768.81		768.81

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Bartik instrument interacts country f ancestry share in 1900 with lagged growth in country f patenting. Controls include city's own patent stock.

business R&D and 70% of venture capital accrues to the top ten patent-producing cities.

Fact 2. *City-level patenting and wages increase with the patent stock in nearby cities.*

The location of innovation matters because its effects on future innovation and production decline with geographic distance. For city j and year t, I run regressions of the form

$$\Delta \log Y_{j,t} = \phi^Y \sum_{i \neq j} \frac{\text{distance}_{ij}^{-1}}{\sum_{i' \neq j} \text{distance}_{i'j}^{-1}} \Delta \log \text{patents}_{i,[t-10,t-1]} + \Xi_{j,t} + v_{j,t}. \tag{2}$$

The outcome Y is the number of patents created in year t or the average wage in year t.

The main independent variable is the total number of patents created in city i in the decade up to t, which I interpret as the stock of neighbors' patents at time t. The regression asks how growth in patenting responds to growth in the patent stock of a city's typical neighbor. Suppose that 3,000 patents were created in Milwaukee in the 1990s and 6,000 in the 2010s. The Milwaukee patent stock would be twice as high in 2010 as in the 2000. The regression compares the growth rate of patenting in nearby Chicago from 2000 to 2010 against the corresponding growth rate in distant Philadelphia.

Columns (1) and (3) of Table 1 present OLS estimates of (2) for patents and average nominal wages, respectively.[3] The controls $\Xi_{j,t}$ include year fixed effects and the local stock of patents from city j, $\log \text{patents}_{j,[t-10,t-1]}$. The coefficient of 1.25 for patents means that a

[3] I stack two decadal changes, 2000 to 2010 and 2010 to 2019. There are a small number of zeros in the patents data, so I approximate the log function with the inverse hyperoblic sine function, which is defined at zero.

1% increase in the stock of neighbors' patents is associated with a 1.25% increase in the annual number of patents.

Instrumental variable

The OLS estimate does not have a causal interpretation if neighboring cities experience a common shock to inventor productivity. For example, nearby cities may specialize in similar industries or have large inter-city migration flows. To isolate a component of neighbors' patent stocks exogenous to local patenting rates, I construct a shift-share Bartik instrument. I build on the demographic variation exploited by Burchardi, Chaney, and Hassan (2019), who find that historical ancestry predicts current foreign direct investment. The instrument interacts the population ancestry of each city in the year 1900 with lagged patent stocks in foreign countries.[4] The predicted change in the patent stock in city i is

$$\Delta \log \widehat{\text{patents}}_{i,[t-10,t-1]} := \sum_f \text{ancestry share}^f_{i,1900} \times \Delta \log \text{patents}^{f-US}_{[t-20,t-11]}. \tag{3}$$

I exclude patents with US authors to avoid any mechanical correlation based on coauthorship links. The instrument is then the distance-weighted average of predicted patent growth, analogous to the independent variable in (2). Two-stage least squares estimates are presented in columns (2) and (4) of Table 1. The 2SLS estimates are slightly larger than OLS, but the difference is not statistically significant. A one standard deviation increase in the average neighbors' patent stock causes a 0.15 standard deviation increase in local patenting and a 0.2 standard deviation increase in local wages.

The logic of the instrument is that inventors in the US are differentially exposed to knowledge spillovers based on the idiosyncratic strength of bilateral connections to foreign countries. Suppose the number of patents produced by German inventors doubles in the 1990s relative to the 1980s. The instrument would predict higher patent growth in 2000 in Milwaukee (with 24% of the population having German ancestry in 1900) relative to Philadelphia (with an 18% German share in 1900), which then implies a higher patent stock in Chicago's average neighbor.

The relevance condition is satisfied if US inventors learn from foreign inventors based on ethnic kinship. Kerr (2008) shows that foreign research scientists in the US share knowledge with their home countries, and Saxenian (1998) describes entrepreneurs and inventors shuttling between California and Asia as important links in the transmission of

[4] Ancestry data are from Fulford, Petkov, and Schiantarelli (2020), who measure population ancestry for US counties.

knowledge. The first-stage is quite strong, with an F statistic near 750 (see Appendix Table A5). In Appendix Table A6 I provide direct evidence on instrument relevance, showing that ancestry shares predict patent citations from city to foreign country.

I appeal to the exogenous shares assumption as laid out in Goldsmith-Pinkham, Sorkin, and Swift (2020) (*cf.* Borusyak, Hull, and Jaravel (2022)). The exclusion restriction is that the ancestry shares of a city's neighbors are unrelated to that city's patent growth except through knowledge spillovers. German ancestry in Milwaukee in 1900 must be uncorrelated with Chicago's patenting in 2000 (conditional on Chicago's patenting in the 1990s). One threat to identification comes from industry shocks. For instance, suppose ethnic Germans have a comparative advantage in chemical manufacturing. Then, Germans and German-Americans patent and work in chemicals. Milwaukee, with its large German population, specializes in chemicals. Chicago develops a chemicals industry because of its proximity to Milwaukee. Suppose further that the chemicals industry experiences a persistent shock to productivity growth over time. My regression would associate Chicago's patent growth with growth in Milwaukee's patent stock. Even in the absence of spatial spillovers, Chicago's patenting would grow because chemists' productivity grew in Germany in the 1990s, in Milwaukee in the 2000s, and in Chicago in 2010. My identification strategy excludes this possibility.

I present several robustness exercises in Appendix A. Throughout, the results change little. Appendix Table A7 accounts for labor demand shocks by controlling for contemporaneous employment growth in each city and in each city's neighbors (distance-weighted). In the same Table I also look at real wages instead of nominal wages by adjusting for local housing prices. In the baseline results I used an arbitrary distance elasticity of -1; Appendix Table A8 shows that the results are qualitatively similar for other values. Appendix Table A9 estimates spatial autocorrelation-robust standard errors developed by Conley (1999) and Müller and Watson (2021). Lastly, one potential concern about the instrument is that ancestry variation picks up differential growth rates in cosmopolitan places. For instance, San Francisco had a relatively high Chinese ancestry share in 1900, and China's patenting grew tremendously. Appendix Table A10 constructs the instrument using only ancestry from Western European countries, and the estimates are virtually unchanged.

In Appendix D I study patenting shocks at the more granular level of occupations within cities. I assign patents to occupations using natural language processing techniques as in Kogan et al. (2021). The regression results look similar.

Fact 3. *Inventors sell patents to nearby firms.*

To rationalize the local productivity effects of patents, I provide new evidence of spatial barriers in the market for technology. Other factors equal, inventors tend to sell their patents to nearby firms.

I use data on inter-firm patent transactions from the USPTO Patent Assignment Dataset (Graham, Marco, and Myers 2018). The USPTO records changes of ownership filed by parties to the patent agreement.[5] Relative to data on patent text and patent citations, the Patent Assignment Dataset has been less widely used in the economics literature.[6] I restrict attention to sales and licensing agreements, and so exclude administrative assignments like employee-to-employer transfers, mergers, name changes, or corrections. Most transactions are sales, and I will henceforth refer to sales and licensing as sales for short. About one quarter of patents granted to US inventors in my data are associated to sales. These patents collect about one third of the total citations made to all patents between 2000 and 2019. In terms of private value—estimated from stock returns by Kogan et al. (2017)—traded patents on average look very similar to other patents. Assigned patents also display a similar geographic distribution to overall patents. The number of assigned patents by city and year has a correlation of 0.93 with the number of total patents by city and year, and the Herfindahl index across cities of the two groups are virtually the same (0.0318 versus 0.323, respectively, in 2019).

As an example, consider patent number 10,433,978. This patent, "Systems and Methods for Adjacent Vertebral Fixation," was granted October 8, 2019 to Dr. Dennis Bullard, a retired spinal surgeon in Raleigh, North Carolina. The patent describes a method to insert a stabilizing prosthesis into a patient's spine. Bullard Spine, LLC sold the patent to Absolute Advantage Medical, a medical device firm based in Southern Pines, North Carolina. Dr. Bullard had no formal employment relationship with Absolute Advantage, but Southern Pines is less than 100 km from Raleigh.

To study the relationship between distance and patent sales more systematically, I compute the number of patents invented in city i and sold to firms in city j between 2000 and 2019. I run gravity regressions of the form

$$\log \text{patent sales}_{ij} = \alpha \text{ same city}_{ij} + \beta \log \text{distance}_{ij} + \Xi_{ij} + \epsilon_{ij} \tag{4}$$

[5] While the submission of changes is voluntary, there is incentive to do so. Unrecorded assignments are *de jure* void (Graham, Marco, and Myers 2018).

[6] Akcigit, Celik, and Greenwood (2016) use the Patent Assignment Dataset to study search and matching frictions in the patents market, while Arque-Castells and Spulber (2022) consider how the scope for technology transfer changes the private returns to R&D. Souza (2022) uses technology licensing by Brazilian firms as a measure of technical change.

Table 2: Patent sales gravity regression

	PPML Log sales			
	(1)	(2)	(3)	(4)
Same city	3.87***	2.28***	1.37***	0.735***
	(0.401)	(0.236)	(0.105)	(0.033)
Same state		1.81***		
		(0.162)		
Log distance			-0.557***	-0.476***
			(0.106)	(0.059)
Log predicted citations				2.61***
				(0.499)
Industry distance				-7.95***
				(2.69)
Origin FE	✓	✓	✓	✓
Destination FE	✓	✓	✓	✓
N	779,972	779,972	779,972	779,972
Pseudo R^2	0.89	0.90	0.90	0.91

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the origin state and destination state levels. Models estimated by psuedo-Poisson Maximum Likelihood. Distance is scaled such that the minimum observed log distance is zero and log own-distance is replaced by zero. Industry distance is Euclidean distance of four-digit NAICS level employment shares. Predicted citations based on correlation of patent technology classes.

where Ξ_{ij} are controls always including origin and destination fixed effects. Table 2 presents the estimates. Column (1) has only the same city dummy and column (2) adds a same state dummy. Log distance enters in column (3) with an elasticity of about minus one-half. In all regressions I set the log of own distance to zero and scale distance such that the smallest observed distance in the data is equal to one. Therefore, the same city dummy is the log difference in sales between the origin city and a hypothetical adjacent city.

If nearby cities specialize in similar industries or use similar technologies, then firms might sell to nearby firms based on common characteristics. In that case the apparent relationship between sales and distance would be spurious. I address this concern with two controls intended to capture the similarity of the city pair. The first is the Euclidean distance between sectoral employment shares at the four-digit NAICS level, defined by

$$\text{industry distance}_{ij} = \sqrt{\sum_{ind} \left(\text{emp. sh.}_{i,ind} - \text{emp. sh.}_{j,ind}\right)^2}.$$

If the distance is zero then the cities have the same sectoral composition. The second control is an index of technological similarity, defined by predicted patent citations. I predict citations from city i to city j by interacting cities' patent classes with the national class-to-class citation probabilities. The formula is

$$\widehat{\text{citations}}_{ij} = \sum_c \sum_l \mathcal{C}^{cl} \times \text{pat. sh.}_i^c \times \text{pat. sh.}_j^l$$

where c, l denote one-digit CPC codes and $\mathcal{C}^{cl}$ is the national fraction of citations made by class c patents given to class l patents.

My preferred specification in column (4) of Table 2 adds the industry and technology controls. I find very similar results, with the distance elasticity still around minus one-half. This estimate implies that the probability of a sale from Denver to Boston is about one-fifth of the probability of a sale from Denver to Colorado Springs.[7]

Appendix Table A1 presents a number of additional specifications. I show that the number of assignments increases with the strength of personal connections between cities: migration, travel, and (most robustly) the relative number of Facebook friendships. These variables are themselves steeply declining in distance, as shown in Appendix Table A4. Second, I examine the sensitivity of the results to potential measurement error. The dataset providers attempt to filter out employee-to-employer assignments, which tend to be strongly declining with distance but which do not reflect interfirm technology transfers. As they acknowledge, the algorithm for identifying these assignments is imperfect (Graham, Marco, and Myers 2018). I restrict patent transactions—which, recall, include licenses and sales—to licenses only. This is because licensing agreements are less likely to include employee-employer transfers. The results are in Appendix Table A2. The relationship between licenses and distance continues to be negative and it is not statistically different from my baseline estimate.

Patent citations

A literature pioneered by Jaffe, Trajtenberg, and Henderson (1993) has documented that patent citations decline with distance, even controlling for the geographic distribution of distinct research fields. That paper uses a case-control approach, in which a cited patent is compared to a random "control" patent in the same field. Here, I use the gravity framework of (4). I record the first citation made to each patent, excluding citations to

[7]The straight line distances from Denver to Boston and Colorado Springs are 2482 km and 93 km, respectively. The relative probability is $\exp(-0.48 * \log(2842/93))$.

Table 3: Patent citations gravity regression

	PPML Log citations			
	(1)	(2)	(3)	(4)
Same city	1.96***	1.25***	0.784***	0.369***
	(0.236)	(0.126)	(0.124)	(0.101)
Same state		0.802***		
		(0.102)		
Log distance			-0.252***	-0.175***
			(0.018)	(0.020)
Industry distance				-5.64***
				(0.891)
Log predicted citations				2.32***
				(0.152)
Citing patent city FE	✓	✓	✓	✓
Cited patent city FE	✓	✓	✓	✓
N	839,972	839,972	839,972	839,972
Pseudo R^2	0.93	0.93	0.93	0.94

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the origin state and destination state levels. Models estimated by psuedo-Poisson Maximum Likelihood. Distance is scaled such that the minimum observed log distance is zero and log own-distance is replaced by zero. Industry distance is Euclidean distance of four-digit NAICS level employment shares. Predicted citations based on correlation of patent technology classes.

patents with the same author or owned by the same firm as the citing patent.[8] I then aggregate first citations to the city pair level over the period 2000 to 2019.

The results are in Table 3. My preferred specification, in column (4), controls for industrial and technological proximity. I find a distance elasticity close to -0.2.[9] Patent citations decline with distance, but less steeply than patent sales. An inventor in Boston is 55% as likely to cite a patent from Denver than is an inventor from Colorado Springs; recall that the relative sales probability was just 20%.

Appendix Table A3 displays additional specifications for the patent citations regression.

[8]Restricting attention to the first citation facilitates a comparison with the sales regression, since each patent typically has no more than one sale event but receives multiple citations.

[9]Figueiredo, Guimarães, and Woodward (2015), using a somewhat different empirical specification, find a distance elasticity of about -0.25. Berkes, Gaetani, and Mestieri (2021) find, as I do, that patent citations decline significantly outside the citing patent's city, but they do not report a distance regression.

2.1 Summary of facts

Patenting is geographically concentrated. Geographic concentration matters because patents confer local productivity benefits to inventors and firms. The local effects of patenting are consistent with a local market for technology and substantiate the presence of spatial barriers to knowledge diffusion. Barriers inhibit adoption by firms and learning by other researchers. I turn to an economic model to formalize how diffusion barriers affect the private and social returns to research and to assess whether the geographic distribution of innovation is efficient.

3 A model of spatial growth

I present a model of semi-endogenous growth with imperfect knowledge diffusion across space. Researchers in each region invent new product varieties, which are the building blocks of what I call "knowledge." Varieties are inputs into goods production. Varieties also increase the productivity of future researchers through an intertemporal spillover.

Knowledge diffusion is imperfect in two ways. On the one hand, sales are subject to spatial barriers. A new variety can be sold only in a random subset of regions, which may depend on distance from the inventing location. On the other hand, learning is also subject to spatial barriers. A new variety can be observed by researchers in a random subset of regions, which is again a function of distance. The horse race between sales barriers and learning barriers determines the private and social returns to research.

The economy is populated by two types of households: researchers, of mass $\overline{S}$, and production workers, of mass $\overline{L}$. Researchers are the substantive decision makers. They can freely choose where to work and I denote by S_i the endogenous number of researchers in region i. By contrast, production employment in region i is exogenously given by L_i, with $\sum_i L_i = \overline{L}$. For short I refer to production employment as "labor." The strong assumption of exogenous labor supply is helpful to fix ideas, as it lets me isolate researchers' decisions. In the quantitative model I relax this assumption by considering mobile, heterogeneous workers.

The state of the economy is summarized by K_i, the cumulative stock of knowledge created by researchers in region i. In the quantitative analysis I will measure K_i using patents. Innovation and diffusion determine how knowledge evolves over time. Time is continuous and indexed by t.

3.1 Model

Production

I begin with the static production part of the model. Each region produces a homogeneous final good assembled from a set of varieties. This good is the economy's numéraire. Since the final good is homogeneous, there will be no trade in equilibrium; in the quantitative model I introduce regional trade in final goods. Let the measure of varieties available in region j at time t be $M_{j,t}$. Output in region j is equal to

$$X_{j,t} = \left(\int_0^{M_{j,t}} x_{j,t}(m)^{\frac{\varepsilon-1}{\varepsilon}} \, dm \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{5}$$

where $\varepsilon > 1$ is the elasticity of substitution between varieties and $x_{j,t}(m)$ is the quantity of variety m used. Innovation and diffusion determine the evolution of $M_{j,t}$, but at every instant firms and workers take $M_{j,t}$ as given.

Each variety is produced by a monopolistically competitive firm. I assume that monopolists can only sell their varieties locally, although later on I show that the insight of the model obtains even when varieties are traded. Each monopolist can transform 1 unit of labor into $A_{j,t}$ units of output. $A_{j,t}$ is an exogenous, region-specific labor productivity term. The monopolist then sells its variety to a competitive final good producer operating the technology (5). Monopolists face constant-elasticity demand and so choose to set a constant markup $\frac{\varepsilon}{\varepsilon-1}$ over marginal cost.

All monopolists within a region have the same marginal costs and face the same demand. They therefore behave identically and choose the same output $x_{j,t}(m) = x_{j,t}$. In equilibrium this output must exhaust the supply of labor, which means each monopolist's output is

$$x_{j,t} = \frac{A_{j,t} L_j}{M_{j,t}}. \tag{6}$$

Substituting this output into the production function (5) yields region j's total output

$$X_{j,t} = A_{j,t} L_j M_{j,t}^{\frac{1}{\varepsilon-1}}. \tag{7}$$

Expanding varieties increase output, disciplined by the elasticity of substitution ε. A low ε means that varieties are relatively poor substitutes and implies high gains from variety.

Flow profits $\Pi_{j,t}$ accrue to each monopolist. It is straightforward to show that monopoly pricing grants profits equal to a share $\frac{1}{\varepsilon}$ of revenue. Aggregate profits $\Pi_{j,t} M_{j,t}$ then command a share $\frac{1}{\varepsilon}$ of value-added $X_{j,t}$. The wages of production workers, $W_{j,t}$, absorb the remaining share. Profits and wages in region j are

$$\Pi_{j,t} = \frac{1}{\varepsilon} A_{j,t} L_{j,t} M_{j,t}^{\frac{2-\varepsilon}{\varepsilon-1}}, \tag{8}$$

$$W_{j,t} = \frac{\varepsilon - 1}{\varepsilon} A_{j,t} M_{j,t}^{\frac{1}{\varepsilon-1}}. \tag{9}$$

Profits increase in local market size $A_{j,t}L_{j,t}$. In the empirically relevant case $\varepsilon \geq 2$, profits per variety decrease in the measure of local varieties. I maintain this assumption throughout the paper. Wages increase in the measure of varieties $M_{j,t}$. I now turn to the research process.

Research

Research labs employ researchers to create new varieties. A lab can hire researchers at competitive wage $R_{i,t}$ and produce under decreasing returns to scale with elasticity $\gamma < 1$. Each lab in i must pay a flow operating cost $F_{i,t}$, which I interpret as the rent for a fixed unit of land as in Grossman and Helpman (2018). Research firms can freely enter and secure zero profits net of rents.[10]

I model the knowledge accumulation process as in Jones (1995). A lab employing s researchers creates a flow of new varieties equal to $Z_{i,t} N_{i,t}^{\lambda} s^{\gamma}$, where $\lambda < 1$ is a parameter. Summing across labs, the local knowledge stock evolves as

$$\dot{K}_{i,t} = Z_{i,t} N_{i,t}^{\lambda} S_{i,t}^{\gamma}. \tag{10}$$

Given an initial condition $K_{i,0}$, the local knowledge stock at time t is the measure of varieties ever invented locally,

$$K_{i,t} = K_{i,0} + \int_0^t \dot{K}_{i,\tau} d\tau. \tag{11}$$

Researchers' productivity depends on two terms: exogenous $Z_{i,t}$ and endogenous $N_{i,t}$. $Z_{i,t}$ is region i's comparative advantage in research. For example, $Z_{i,t}$ may reflect universities, scientific organizations, government labs, or other durable research inputs. $N_{i,t}$ the knowledge pool in region i. The knowledge pool represents an externality in the form of

[10] Decreasing returns introduce local congestion in research and avoids a counterfactual "black hole" in which all research concentrates in a single region. In Appendix B.1 I provide two isomorphisms with constant returns to scale and no fixed costs. In the first, workers have heterogeneous research ability across regions. In the second, research uses local intermediate goods supplied subject to increasing marginal cost.

intertemporal knowledge spillovers as in Romer (1990) and Jones (1995). $N_{i,t}$ depends on the stocks of local knowledge and knowledge from other regions, and I will elaborate on its structure shortly. The parameter λ is the intertemporal spillover elasticity. A higher λ means research output is more sensitive to spillovers, $\lambda = 0$ corresponds to the case without spillovers, and a negative λ means that knowledge accumulation is crowded out by existing knowledge. I assume $\lambda > 0$ throughout, which will be the empirically relevant case. The intertemporal spillover is external. Research labs take $N_{i,t}$ as given and ignore their infinitesimal contribution to it. Later I discuss a model extension in which a large lab is able to partly internalize the spillover.

The knowledge pool is shaped by diffusion barriers. Researchers in region i only learn about a fraction $\delta^K_{ji} \in [0,1]$ of the knowledge ever created in region j (including i). The knowledge pool is then

$$N_{i,t} = \sum_j \delta^K_{ji} K_{j,t}. \tag{12}$$

The δ^K_{ji} are parameters I interpret as spatial barriers to learning, for instance because exchanging information is easier face-to-face. $\delta^K_{ji} < 1$ means that diffusion from j to i is incomplete in that researchers in i do not fully absorb knowledge from j.[11]

The model features semi-endogenous growth because the intertemporal spillover λ is strictly less than one. Economic growth in this class of models ceases asymptotically unless there is sustained growth in research inputs. I therefore assume $Z_{i,t}$ grows at constant rate g^Z in every region and at every point in time.

In principle, the measure of varieties supplied in region j, $M_{j,t}$, could be equal to the grand sum of knowledge created anywhere in the world and at any time in the past. However, analogous to imperfect learning, only a subset of region i's knowledge stock $K_{i,t}$ translates into active varieties in other regions j. The set of varieties actually sold to and adopted by firms is limited by spatial barriers in the sales process, which I turn to presently.

Technology transfer

Once a lab creates a variety, it licenses the intellectual property to monopolists. I study a franchise model in which the lab makes non-exclusive sales to different producers in different regions. The licensing process is subject to frictions. When a lab in region i creates an idea, it draws a set of once-and-for-all diffusion shocks, one for each region j (including i). Diffusion shocks are binary: for each j, the lab can either license the idea to a

[11] My parametrization of incomplete knowledge diffusion follows Peri (2005).

producer in j, or it cannot. I assume that the diffusion shocks are drawn independently across destinations.

A lab in i can license its idea to producers in j with probability δ_{ij}^X. The parameters $\delta_{ij}^X \in [0,1]$ represent barriers to selling, analogous to the barriers to learning in δ_{ij}^K. At this point I make no assumption on the relationship between δ_{ij}^X and δ_{ij}^K. I think of the δ_{ij}^X as a reduced-form way of capturing frictions in the process of getting a product to market. For example, product adoption might require marketing or familiarization which are easier face-to-face. Barriers to selling generate differences in market size. For example, a lab in Austin may frequently licenses its product to firms in Dallas, El Paso, and Houston; less frequently in Chicago and Detroit; and less frequently still in New York and Boston. Once an idea diffuses to a region, a fringe of potential producers compete á la Bertrand to license it. The winner receives a local monopoly to fabricate the variety. By Bertrand competition, the lab receives the full flow of profits in perpetuity.

I note three strong assumptions in my model of technology transfer. First, patents never expire. This is for simplicity, and is relaxed in the quantitative model. Second, monopolies are local. A monopolist in El Paso does not compete with a monopolist in Boston producing the same product. Allowing monopolists to trade varieties across space would make the model intractable because profits would depend on the set of producers, which in turn would be determined by the realization of the diffusion shocks. That problem would have combinatorial complexity because the set of potential competitors is the power set of locations.[12] Third, I do not explicitly model "patent trolls," firms which buy patents not for the sake of production but instead to pursue frivolous litigation claiming infringement. My model's market structure ensures researchers extract the full profits created by their ideas, and since patents are costlessly enforced, trolling is fruitless.

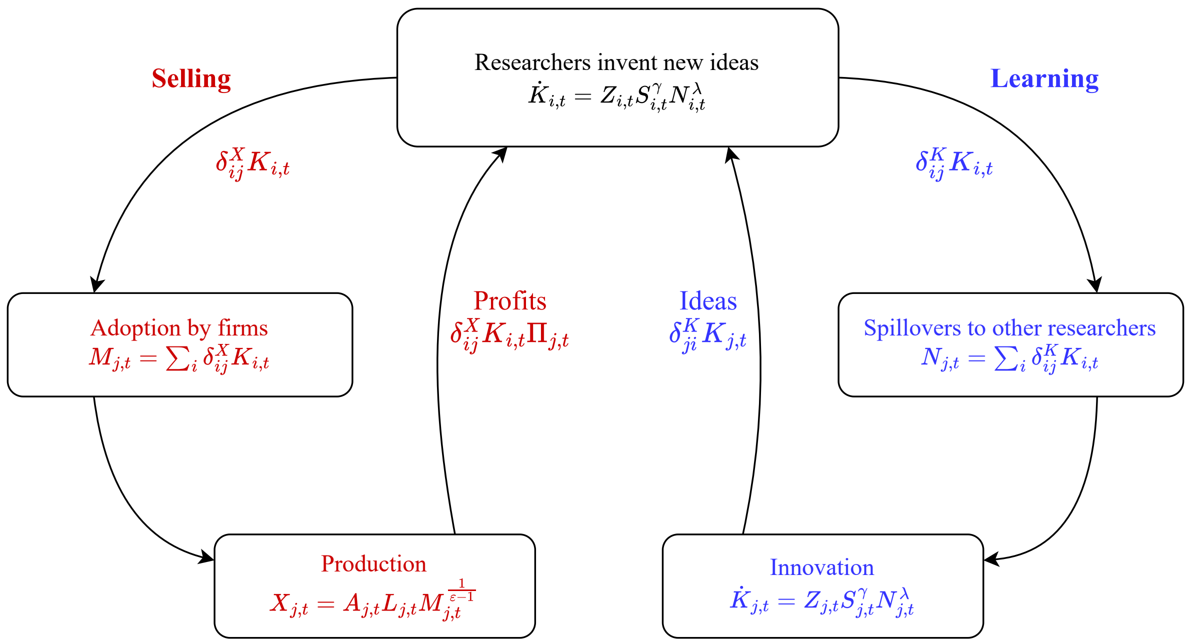
The technology transfer process admits a simple formula for the measure of varieties in each region. The measure of varieties in region j is simply the subset of knowledge available from every region i.

$$M_{j,t} = \sum_i \delta_{ij}^X K_{i,t}. \tag{13}$$

Note the symmetry between the knowledge pool (12) and the measure of varieties (13). The

[12]Instead of licensing intellectual property, I could have instead considered a model in which inventors produce their varieties at home and then export the output to other regions. The reason I focus on licensing instead of exporting is so that innovation is not constrained by market size. Silicon Valley's innovation share is an order of magnitude higher than its production share. The observed distribution of production is not consistent with a model in which Silicon Valley firms produce and export their goods and services directly to the rest of the country. However, as I discuss in Appendix B.7.1, a trade-only model delivers similar insights to my licensing model.

Figure 2: Model schema



knowledge pool goes into producing more knowledge, while the measure of varieties goes into producing goods. Both variables are linear combinations of the local knowledge stocks from *every* region, $\{K_{1,t}, K_{2,t}, K_{3,t}, \ldots\}$. The weights are given, respectively, by learning barriers δ^K_{ij} and selling barriers δ^X_{ij}. Having separate parameters δ^X_{ij} and δ^K_{ij} enforces a conceptual distinction between selling barriers and learning barriers. Selling barriers are internalized by profit-maximizing researchers while learning barriers govern an externality and so are ignored. It may seem natural that $\delta^K_{ij} < \delta^X_{ij}$ to the extent that new knowledge has a tacit component or if innovation arises from serendipitous encounters.[13] On the other hand, researchers may be likelier than managers to travel or collaborate, consistent with $\delta^K_{ij} > \delta^X_{ij}$. The knife-edge case $\delta^K_{ij} = \delta^X_{ij}$ is possible too. I will discipline these parameters with the data on patent sales and citations when I quantify the model.

Figure 2 represents the basic structure of technology transfer and knowledge spillovers.

Investment

The lifetime utility of a household at time t with a consumption path $[c_\tau]^\infty_{\tau=t}$ is

$$u_t = \int_t^\infty \exp\left(-\rho\left(\tau - t\right)\right) \log c_\tau d\tau. \tag{14}$$

[13] At face value, this configuration would imply that a region can produce a variety that it does not know about. My stance is that the ability to use a technology in production is different from the ability to use it to design a better technology.

I assume that researchers and workers are hand-to-mouth. Labs finance research by issuing equity to immobile local capitalists. Capitalists own the stock of land, the quantity of which I normalize to unity, and rent it to labs. Capitalists have the same intertemporal preferences as households, and so given a rate of consumption growth g_t, the dividend yield ι_t is pinned down by the Euler equation

$$g_t = \iota_t - \rho. \tag{15}$$

Equilibrium

I close the model with the demand for and supply of research labor. Labs hire researchers to maximize profits, while researchers choose the region in which to work to maximize utility.

The unit revenue facing a lab in region i at time t is the value of a variety created there. The value is the total discounted expected profits, where the expectation is taken with respect to the diffusion shocks. The value takes the form

$$V_{i,t} = \sum_j \delta_{ij}^X \int_t^\infty \exp\left(-\int_t^\tau \iota_{\tau'} d\tau'\right) \Pi_{j,\tau} d\tau. \tag{16}$$

A typical lab in region i then solves

$$\max_s V_{i,t} Z_{i,t} N_{i,t}^\lambda s^\gamma - R_{i,t} s - F_{i,t}, \tag{17}$$

ignoring its own infinitesimal effect on the future knowledge pool. The labs' first order condition delivers the regional demand curve for research labor,

$$R_{i,t} = \gamma V_{i,t} Z_{i,t} N_{i,t}^\lambda S_{i,t}^{\gamma-1}. \tag{18}$$

The land rent $F_{i,t}$ adjusts so free entry in research binds.

Researchers are freely mobile across space and choose to maximize current earnings $R_{i,t}$. In an equilibrium with positive research employment everywhere, the prevailing research wage is equalized across space,

$$R_{i,t} = R_t. \tag{19}$$

The market for research labor clears at every instant,

$$\sum_i S_{i,t} = \overline{S}. \tag{20}$$

Aggregate income is the sum of wages to production workers, wages to researchers, profits, and rents. Aggregate output consists of the final good, produced in each region according to (5), and research services, produced according to (10). The final good is consumed and research services are invested. Because the final good $X_t = \sum_j X_{j,t}$ is used only for consumption, goods output growth g_t^X equals consumption growth g_t at every point in time.

I use bold-face symbols to denote vectors. An equilibrium is a path of consumption growth g_t consistent with the evolution of final goods output (5); profits $\mathbf{\Pi}_t$ given by (8); worker wages $\boldsymbol{W}_t$ given by (9); knowledge stocks $\boldsymbol{K}_t$, knowledge pools $\boldsymbol{N}_t$, and production varieties $\boldsymbol{M}_t$ consistent with (11), (10), (12), and (13); a path of interest rates ι_t given by the Euler equation (15); variety values $\boldsymbol{V}_t$ given by (16); and researcher allocations $\boldsymbol{S}_t$ and wages $\boldsymbol{R}_t$ consistent with the demand (17) and supply (19) of research labor. Researchers choose to maximize their wage at every instant. But innovation is forward-looking because the research wage encodes the full flow of profits generated by innovation.

Balanced growth path

I specialize to a balanced growth path (BGP) equilibrium. On a balanced growth path, the researcher allocation and interest rate are constant. From knowledge accumulation (10) and the knowledge spillover (12), knowledge grows at exogenous rate $g^K = \frac{g^Z}{1-\lambda}$. From output (5) and the measure of varieties (13), goods output grows at rate $g = \frac{g^K}{\varepsilon-1}$. Aggregate income, consumption, rents, and aggregate profits also grow at rate g. The flow of new knowledge, $\dot{K}_{i,t}$, is equal to $g^K K_{i,t}$ by virtue of exponential growth. The model features semi-endogenous growth, so the growth rate of output and consumption is exogenous in the long run. However, the levels of output and consumption are endogenous to the allocation of researchers.

In Appendix B.2, I derive the growth rates of all endogenous variables. I also show that the BGP is asymptotically independent of the initial knowledge stocks $K_{i,0}$.

3.2 Decentralized equilibrium

The allocation of researchers, $\boldsymbol{S}$, is the economy's key endogenous variable. The research allocation pins down the full path of knowledge accumulation, which then determines consumption, wages, and profits. The equilibrium numbers of researchers and knowledge stock in every location are jointly characterized by two equations, knowledge accumulation (10) and labor demand (18). Rearranging these two equations gives

$$S_{i,t} = \varpi^S \left(V_{i,t} N_{i,t}^{\lambda} Z_{i,t} \right)^{\frac{1}{1-\gamma}} \tag{21}$$

$$K_{i,t} = \varpi_t^K \left(V_{i,t}^{\gamma} N_{i,t}^{\lambda} Z_{i,t} \right)^{\frac{1}{1-\gamma}} \tag{22}$$

where ϖ^S and ϖ_t^K are uninteresting equilibrium constants. The value is the present discounted value of profits a research lab in region i can expect to receive from its variety, accounting for outward selling barriers. The value is

$$V_{i,t} = \frac{1}{\rho + g^K} \sum_j \delta_{ij}^X \Pi_{j,t}, \tag{23}$$

which I obtain by integrating (16) using the interest rate implied by the Euler equation. The values $V_{i,t}$ and knowledge pools $N_{i,t}$ are endogenous but pinned down by $K_{i,t}$.

Conditions (21) and (22) say that the number of researchers $S_{i,t}$ and local knowledge stock $K_{i,t}$ both increase with the private variety value $V_{i,t}$ and local research productivity $N_{i,t}^{\lambda} Z_{i,t}$. This provides a natural explanation as to why so many patents are created in Boston and Silicon Valley. The value of each variety is high because researchers there are close to large, productive markets. Research productivity is high because of a combination of a deep learning pool and high exogenous productivity.

While some regions are better places to innovate on average, they are all equally attractive on the margin, thanks to free mobility and local congestion.[14] Rearranging the labor demand curve, the research wage is just the number of new varieties created by the marginal researcher, multiplied by the value of each of those varieties:

$$\underbrace{R_t}_{\substack{\text{research}\\ \text{wage}}} \propto \underbrace{K_{i,t}/S_{i,t}}_{\substack{\text{\# varieties}\\ \text{per}\\ \text{marginal}\\ \text{researcher}}} \times \underbrace{V_{i,t}}_{\substack{\text{value}\\ \text{per}\\ \text{variety}}}. \tag{24}$$

[14] Congestion is internal to each research lab. In Appendix B.7.2 I show that allowing local agglomeration externalities in the number of researchers does not change the implications of the model.

A given number of researchers can be equivalently compensated by a large flow of new varieties, each with a low value, or by a small flow of new varieties, each with a high value. Previewing the source of misallocation in this economy, a social planner cares about more than profits. Varieties that generate low profits can still deepen the knowledge pool, and they can be produced at lower cost because those locations have lower congestion in research.

Inequality

Incomplete diffusion generates spatial inequality because some regions adopt more varieties than others. The level of production wages is given by (9). In terms of knowledge stocks, production wages are

$$W_{i,t} = \frac{\varepsilon - 1}{\varepsilon} A_{j,t} \left(\underbrace{\sum_{i'} \delta^X_{i'i} K_{i',t}}_{M_{i,t}} \right)^{\frac{1}{\varepsilon-1}} .$$

Wages are increasing in the stock of available varieties, which are in turn determined by the region's proximity to researchers through the barriers δ^X. The variance of log wages is given by

$$\text{Var}\left(\log W_{i,t}\right) = \text{Var}\left(\log A_{i,t}\right) + \frac{1}{\left(\varepsilon - 1\right)^2}\text{Var}\left(\log M_{i,t}\right) + 2\frac{1}{\varepsilon - 1}\text{Cov}\left(\log A_{i,t}, \log M_{i,t}\right) \tag{25}$$

Diffusion barriers affect wage dispersion through two channels. First, wage dispersion increases mechanically when $M_{i,t}$ differs across space, as indicated by the second term on the right-hand side of (25). Second, wage dispersion rises to the extent that the set of varieties $M_{i,t}$ covaries positively with the level of exogenous productivity $A_{i,t}$, as in the third term in (25). Because profits are increasing in $A_{i,t}$ (recall (8)), innovation tends to be directed toward productive regions and so $M_{i,t}$ and $A_{i,t}$ do tend to covary positively, all else equal.

Frictionless benchmark

The model nests the special case of no spatial barriers in knowledge diffusion when $\delta^X_{ij} = \delta^K_{ij} = 1$ for all i and j. Then, firms can buy and researchers can learn from every variety in the economy, independent of its origin.

The implications of the frictionless benchmark are stark. In that case, the allocation of researchers is determined only by exogenous fundamentals $\boldsymbol{Z}_t$, since values $\boldsymbol{Z}_t$ and the knowledge pools $\boldsymbol{N}_t$ do not vary by location. Knowledge spillovers operate but are absorbed into a constant. From (22), the stock of local knowledge, $K_{i,t}$, is exactly proportional to the number of researchers, hence to the exogenous fundamental $Z_{i,t}^{1-\gamma}$. In this version of the world, Boston produces many patents because it just happens to be very good at innovation.

Meanwhile, the wage for production workers, (9), is equal to $W_{i,t} = \frac{\varepsilon-1}{\varepsilon} A_{i,t} M_t^{\frac{1}{\varepsilon-1}}$, where M_t is the common stock of varieties. Wage differences across regions arise purely from differences in exogenous fundamentals $\boldsymbol{A}_t$, since firms use the same varieties everywhere.

4 Research policy

I now study the efficiency properties of the decentralized equilibrium. As in standard models of endogenous and semi-endogenous growth, my model features an externality: researchers are not compensated for the spillovers they impart to future researchers. The typical policy prescription is to draw labor out of production and into research (for example, by subsidizing research employment). By fixing the number of researchers at $\overline{S}$, I deliberately assume away misallocation between production and research. Instead, my model describes a new spatial margin of research misallocation.

Researchers are misallocated across space to the extent that the private returns to research, encoded in the market value of a new variety, diverge from the social returns to research, encoded in the spillover value of a new variety. Private returns depend on profits. Regions with low selling barriers to large or productive markets are profitable platforms for innovation. By contrast, social returns emerge from opportunities for learning. Regions with low learning barriers to research-intensive regions are fruitful platforms for innovation. Spatial mismatch between profits and spillovers therefore generates spatial misallocation.

4.1 Reallocation

I build intuition for the mechanics and effects of knowledge diffusion by showing what happens to innovation and wages following a small shock to the research allocation. In Section 4.2 I move to the social planner's problem of choosing the optimal allocation of researchers to maximize aggregate consumption.

Define the learning absorption matrix Ω^K with typical entry $\Omega^K_{ij} = \frac{\delta^K_{ij} K_i}{N_j}$. This matrix records the share of region j's knowledge pool originating from region i. Similarly, define the usage absorption matrix Ω^X with typical entry $\Omega^X_{ij} = \frac{\delta^X_{ij} K_i}{M_j}$, which records the share of region j's technology originating from region i.

Lemma 1 (Reallocation shock)**.** *Consider a policy which induces a small reallocation of researchers, $d \log \boldsymbol{S}$. The change in knowledge stocks from old BGP to new is*

$$d \log \boldsymbol{K} = \gamma \left(I - \lambda \Omega^{K'} \right)^{-1} d \log \boldsymbol{S}. \tag{26}$$

The change in wages is

$$d \log \boldsymbol{W} = \frac{1}{\varepsilon - 1} \Omega^{X'} d \log \boldsymbol{K}. \tag{27}$$

Proof. See Appendix B.3. □

Equation (26) traces the propagation of knowledge spillovers through space. Knowledge in i spills over to i', which in turns spills over to i'' and back to i, and so forth. The spillovers matrix $\left(I - \lambda \Omega^{K'} \right)^{-1}$ can be expressed in the equivalent form

$$\left(I - \lambda \Omega^{K'} \right)^{-1} = I + \lambda \Omega^{K'} + \lambda^2 \Omega^{K'^2} + \ldots \tag{28}$$

which evokes a feedback loop: the direct effect, the first round of spillovers, second round, etc.[15] In turn, (27) translates the shock into wage changes, modulated by the usage absorption matrix Ω^X. My theory draws heavily on Liu and Ma (2021), who study an innovation network of multiple sectors. That paper regards particular sectors as "upstream" or "downstream" in the innovation network (for example, advanced manufacturing may have a higher spillover than consumer goods). Here, the innovation network is mediated by spatial barriers to the diffusion of ideas.

In the benchmark case of full diffusion, reallocating researchers has zero effect on output and wages.

Proposition 1 (Reallocation and wages)**.** *Suppose there were no spatial barriers to learning or selling: $\delta^K_{ij} = \delta^X_{ij} = 1 \; \forall \; i, j$. Following a small reallocation of researchers $d \log \boldsymbol{S}$, the change in knowledge stocks is determined only by the direct effect: $d \log \boldsymbol{K} = \gamma d \log \boldsymbol{S}$. Output and wages do not change anywhere: $d \log \boldsymbol{X} = d \log \boldsymbol{W} = 0$.*

[15] By construction the matrix $\lambda \Omega^{K'}$ has a spectral radius less than one (Ω is right stochastic), hence is convergent.

Proof. See Appendix B.4. □

The intuition behind the first part of Proposition 1 is as follows. When there are no spatial barriers in selling, then the variety value V_t is equalized across space. From (24), free mobility guarantees that the number of varieties per researcher, $K_{i,t}/S_{i,t}$, is equalized too. This is equivalent to saying that the marginal physical product of research labor is equalized in every region, because the average product and marginal product are proportional for a constant-elasticity function. In turn, equalizing the marginal physical product of research labor across regions maximizes the aggregate stock of knowledge $\sum_i K_{i,t}$, taking the size of the spillover as given. But when there are no spatial barriers in learning, then the spillover itself coincides with the aggregate stock of knowledge, $N_t = \sum_i K_{i,t}$. Envelope-style reasoning then guarantees that any deviation from this maximum point yields no increase in N_t. Reallocation can increase knowledge in the recipient region, but only at the expense of knowledge in the sending region, such that aggregate knowledge cannot increase.

The second part of Proposition 1 arises because output and wages are simply a function of $M_t = N_t = \sum_i K_{i,t}$ when there are no barriers to selling. If reallocation does not change the aggregate stock of knowledge, then it does not change output or wages. Maximizing the aggregate stock of knowledge is optimal because all workers benefit equally from this common aggregate stock.

Proposition 1 shows that it is hard to justify place-based R&D policy unless there are spatial barriers to knowledge diffusion. The externality brought about by intertemporal knowledge spillovers is operative even in the full-diffusion benchmark, but it is of the same size everywhere and so is not distorted by researchers' choices in equilibrium.

4.2 Planner's problem

I now present the paper's main theoretical result. I study the optimal control problem of a benevolent social planner who chooses a research allocation $\boldsymbol{S}_t$ to maximize the present value of log aggregate consumption, given the distribution of production employment.[16]

[16] Spatial models commonly assume free mobility with compensating differentials. Welfare is then the level of expected utility in the economy, typically a power mean of wages with some amenity weights. The level of welfare in these models depends on the functional form of households' preference shocks, e.g. Fréchet. However, the functional form is not identified from location choice data, complicating the interpretation of expected welfare (Davis and Gregory 2022). I choose to maximize consumption because the problem is well-posed without any assumptions on labor supply.

The problem is

$$\max_{S_{i,\tau}} U_t = \int_t^{\infty} \exp\left(-\rho\left(\tau - t\right)\right) \log\left(\sum_j X_{j,t}\right) d\tau, \tag{29}$$

subject to the law of motion for knowledge stocks. In practice, a policymaker may have goals other than maximizing aggregate consumption—for example, reducing spatial inequality. Such a redistributive motive could be addressed through reallocation of researchers, since wages are increasing in the measure of available varieties from (9). I return to this point in Section 4.4.

The planner internalizes knowledge spillovers. Adding researchers boosts knowledge accumulation directly in the receiving region—the first term in the sum (28)—and indirectly in regions with a high learning absorption share—the higher-order terms in (28). By manipulating the research allocation, the planner can engineer the distribution of knowledge stocks according to (26). The target distribution of knowledge stocks accounts for the extent to which knowledge is actually useful in production, which shows up in the usage absorption process (27).

Proposition 2 (Place-based policy)**.** *On the BGP, the optimal research allocation is*

$$\frac{S_i^*}{\overline{S}} = K_{i,t}^* \left\{ \frac{\rho/g^K + (1-\lambda)}{\rho/g^K + 1} \underbrace{\sum_j \frac{\delta_{ij}^X}{M_{j,t}^*} \frac{X_{j,t}^*}{X_t^*}}_{\textit{Market access}} + \frac{\lambda}{\rho/g^K + 1} \underbrace{\sum_j \frac{\delta_{ij}^K}{N_{j,t}^*} \frac{S_j^*}{\overline{S}}}_{\textit{Knowledge spillover}} \right\}. \tag{30}$$

By comparison, the decentralized research allocation is

$$\frac{S_i}{\overline{S}} = K_i \sum_j \frac{\delta_{ij}^X}{M_{j,t}} \frac{X_{j,t}}{X_t}. \tag{31}$$

In equivalent matrix form, denoting research and production shares $S_{i,t} \equiv S_{i,t}/\overline{S}$, $x_{i,t} \equiv X_{i,t}/X_t$,

$$\begin{aligned} \boldsymbol{s}^* &= \left(I - \frac{\lambda}{\rho/g^K + 1}\Omega^{K*}\right)^{-1} \frac{\rho/g^K + (1-\lambda)}{\rho/g^K + 1}\Omega^{X*}\boldsymbol{x}^* \\ \boldsymbol{s} &= \Omega^X \boldsymbol{x}. \end{aligned}$$

Proof. See Appendix B.5. The decentralized allocation is an algebraic re-arrangement of (22). □

Market access encodes profits, since by the monopolistic competition structure aggregate profits are a constant share of output. The knowledge spillover in a region corresponds to the number of researchers who can learn from research conducted there. In the optimal allocation, the planner balances market access against knowledge spillovers. In the decentralized allocation, researchers follow market access alone. If the discount rate ρ is high relative to the growth rate g^K, or if $\lambda \to 0$ such that spillovers vanish, then the optimal allocation approaches the decentralized allocation.

The role of λ highlights that learning is the essential source of inefficiency. Without learning, the research allocation is efficient even when diffusion is imperfect. Selling barriers are of course undesirable and consumption would be higher without them. But the planner cannot overcome these barriers any better than private researchers.

Finally, the planner's allocation coincides with the decentralized allocation in the frictionless benchmark. In that case, the right-hand sides of (30) and (31) do not depend on i, and both collapse to the condition that the number of researchers S_i is proportional to the knowledge stock $K_{i,t}$. This condition admits from the demand curve for research labor, as described earlier.

4.3 Policy implementation

The optimal allocation can be implemented by a region-specific subsidy or tax to research employment. Define the variable Θ_i as region i's knowledge spillover relative to its market access,

$$\Theta_i \equiv \frac{\sum_j \frac{\delta^K_{ij}}{N^*_{j,t}} \frac{S^*_{j,t}}{S}}{\sum_j \frac{\delta^X_{ij}}{M^*_{j,t}} \frac{X^*_{j,t}}{X^*_t}} = \frac{\Omega^{K*} \boldsymbol{s}^*}{\Omega^{X*} \boldsymbol{x}^*} = \frac{\text{Knowledge spillover}_i}{\text{Market access}_i}. \tag{32}$$

Θ_i is a sufficient statistic for the optimal policy rule.

Corollary 1 (Place-based subsidy)**.** *The planner's allocation* (30) *can be implemented by a region-specific subsidy rate to research employment, ς_i. The subsidy is*

$$1 + \varsigma_i \propto 1 + \frac{\lambda}{\rho / g^K + 1 - \lambda} \Theta_i, \tag{33}$$

where the constant of proportionality can be chosen to balance the budget.

Proof. See Appendix B.6. □

It is not sufficient for the planner to subsidize research in high-spillover locations. Silicon Valley may generate high spillovers, but it presumably also generates high profits (which is why there are so many researchers there to begin with). Rather, the planner targets regions which under-innovate in equilibrium. These are the regions in which spillovers are high relative to market access. For example, a region with poor market access—say because it is remote from large population centers—will tend to innovate little, all else equal. If this region's ideas are relatively accessible to other researchers, though, then its innovation is too low. Conversely, suppose there were a region with good market access, perhaps because it is itself large, but which is relatively inaccessible to researchers. In equilibrium this region would innovate excessively because few researchers could learn from the ideas created there.

The following special cases formalize this intuition.

1. (No barriers) In the frictionless benchmark with $\delta_{ij}^K = \delta_{ij}^X = 1$, knowledge spillovers and market access are each equalized across all regions. Then the wedge between private and social values does not vary across space and the optimal subsidy (33) is zero.

2. (No selling barriers) Let $\delta_{ij}^K < 1$ and $\delta_{ij}^X = 1$. All regions have the same market access and are equally profitable for researchers, but some regions have higher knowledge spillovers. The policy (33) subsidizes research in places that already innovate intensively in equilibrium.

3. (No learning barriers) Let $\delta_{ij}^X < 1$ and $\delta_{ij}^K = 1$. Some regions have worse market access and are less profitable for researchers, but all regions have the same knowledge spillover. The policy (33) subsidizes research in places with low market access—places that innovate little in equilibrium. These locations have low-value ideas, so free mobility in research implies that the marginal researcher must be productive there. The planner exploits this by sending more researchers these locations.

4. (No learning) Let $\lambda = 0$. There are no externalities. Selling barriers, if they exist, are internalized by researchers.

5. (No research specialization) Let $\delta_{ij}^K = \delta_{ij}^X$, not necessarily equal to unity, so that knowledge is the same in research and production, $N_j = M_j$. Consider an economy with fundamentals such that all regions are equally specialize in research equally in equilibrium. In such an economy, research earnings RS_j are proportional to labor earnings $W_j L_j$, which are in turn always proportional to output X_j. Then, the policy (33) prescribes an optimal subsidy of zero. In such an equilibrium, demand-side

incentives (which matter for firms) coincide with supply-side incentives (which matter for the planner). This is essentially the one-region model of Romer (1990) and Jones (1995) with a fixed supply of researchers.

4.4 Inequality

Policymakers care about spatial inequality, which the utilitarian planner ignored. Reallocating researchers away from a region may hurt workers who enjoyed privileged access to knowledge formerly created there, and higher innovation elsewhere may only partly offset the direct loss for this group of workers.

It is not clear *ex-ante* whether research policy decreases or increases wage inequality across cities. To see this, it is simplest to reason again in terms of the subsidy (33). Compare two regions 1 and 2 such that

$$\Theta_1 < \Theta_2 \tag{34}$$

The policy dictates a transfer of researchers from region 1 to region 2.

To fix ideas, assume that differences in market access are mostly about local size (say, because selling barriers to other regions are very binding). First, suppose that region 1 has higher exogenous productivity than region 2 but that knowledge spillovers are comparable. Region 2 gains researchers and, because region 2 had lower initial wages, spatial wage inequality falls. Conversely, suppose that region 2 has slightly higher exogenous productivity than region 1, but much a higher knowledge spillover. Region 2 had higher initial wages, so spatial inequality rises as region 2 gains researchers.

In sum, the utilitarian policy increases spatial inequality when large, high-wage cities under-innovate and decreases spatial inequality when small, low-wage cities under-innovate. The former case is likelier when knowledge spillovers are more constrained by distance, while the latter case is likelier when technology adoption is more constrained by distance. In Section 2 I showed that selling is more constrained by distance, which suggests that small, low-wage cities under-innovate in equilibrium. The quantitative model in the next section formalizes this insight.

4.5 Robustness and extensions

Before the quantification, I discuss how alternative modelling choices would affect my results.

In Appendix B.7.1 I replace licensing with trade. The resulting model is similar to Grossman and Helpman (2018) except with mobile researchers. Researchers double as entrepreneurs: they produce the new variety themselves at home and export to other regions. Trade costs act like selling barriers they reduce the profits available to researchers in small, remote regions. A deeper, and perhaps more surprising, implication is that profits are smaller in small regions even under free trade. The reason is that entrepreneurs in small regions are constrained by the size of their local labor pool. They cannot operate at the same scale as entrepreneurs in large markets.[17] Trade in goods is a substitute for trade in ideas but is more tightly bounded by market size.

Appendix B.7.2 returns to the baseline model and considers static agglomeration in research as in Moretti (2021). In that formulation, there is a local externality in the number of researchers, captured by modelling $Z_{i,t}$ as an increasing constant-elasticity function of $S_{i,t}$. It turns out that local agglomeration does not change the results. In particular, the decentralized equilibrium is still efficient without spatial barriers, and the planner's condition (30) continues to hold. While surprising at first glance, the intuition carries over from static spatial models without regional transfers (Kline and Moretti 2014; Fajgelbaum and Gaubert 2020). When the spillover is constant-elasticity, gains from agglomeration in one location are exactly offset by losses from agglomeration in another location.

The baseline model treats researchers as atomistic agents who ignore their effect on the knowledge pool. Appendix B.7.3 studies the problem of a large research lab which can partially internalize knowledge spillovers. The lab mimics the planner's strategy, and its ability to effectuate the planner's solution depends on its size. In practice, the knowledge shares of individual labs are low. The single largest patenting organization in 2019 was IBM, which received 9,253 US patents. While non-negligible, this represents 2.6% of the 354,430 patents granted in the US in 2019 (USPTO 2019).

5 Quantitative model

To implement the optimal R&D policy, I enrich the model from Section 3. First, I introduce regional trade and trade costs, which captures important variation across locations in the private returns to research. Second, I consider heterogeneous worker skill levels, which allows me to discuss between-group inequality. Third, I allow workers to be mobile across locations, which permits employment changes in response to policy. Fourth, I model compensating differentials for researchers in the form of local amenities, which generates additional differences in the private returns to research across space. Fifth, I assume a

[17] Walsh (2019) documents that establishments in larger cities are larger throughout the life cycle.

more realistic life cycle for new varieties. Instead of one-shot diffusion with perpetual patent protection, I allow varieties to diffuse at a constant hazard rate and be imitated at a constant hazard rate. An imitated variety no longer generates profits. Derivations are in Appendix B.8.

5.1 Additional features in the quantitative model

Trade

Each region produces a differentiated good á la Armington (1969). Let the output price of region i's good be $P_{i,t}$, and assume iceberg trade costs $\varkappa_{ij} \leq 1$ such that the unit price of i's variety paid in j is $P_{ij,t} := P_{i,t}/\varkappa_{ij}$. Regional goods are substitutes in consumption with constant elasticity of substitution $\varphi > 1$. I denote the price index in region j by $\mathbb{P}_{j,t} := \left(\sum_i P_{ij,t}^{1-\varphi}\right)^{1/(1-\varphi)}$ and the share of region j's spending dedicated to region i's good by $\xi_{ij,t} = \left(\frac{P_{ij,t}}{\mathbb{P}_{j,t}}\right)^{1-\varphi}$.

Production function

Workers are heterogeneous with respect to their type q. There are two worker types, skilled ($q = h$) and unskilled ($q = \ell$). The exogenous mass of each type is $\overline{L}_q$. I maintain the assumption that aggregate research employment is fixed at $\overline{S}$.

Following the literature on directed technical change (Acemoglu 1998; Acemoglu and Zilibotti 2001), I assume that production combines intermediate varieties specific to either skilled or unskilled labor. Knowledge is therefore appropriate for specific skill groups. Given labor supplies $L_{j,q,t}$ and effective measures of skill-specific varieties $\mathbb{M}_{j,q,t}$, regional output is

$$X_{j,t} = \left(\left(A_{j,\ell,t}L_{j,\ell,t}\mathbb{M}_{j,\ell,t}^{\frac{1}{\varepsilon-1}}\right)^{\frac{\sigma-1}{\sigma}} + \left(A_{j,h,t}L_{j,h,t}\mathbb{M}_{j,h,t}^{\frac{1}{\varepsilon-1}}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}. \quad (35)$$

Each type-q worker in region j collects a wage $W_{j,q,t}$. The effective measure of skill-specific varieties, $\mathbb{M}_{j,q,t}$, is a scalar multiple of the actual measure $M_{j,q,t}$ when there is imitation, a point I return to later.

Worker mobility

Workers can move in response to real wages and amenities. Let the exogenous amenity enjoyed by households of type q in region j be $B_{j,q}$. I follow Allen and Arkolakis (2014) and

Redding and Rossi-Hansberg (2017) and assume that regional employment for type q is

$$L_{j,q,t} = \frac{B_{j,q}\left(W_{j,q,t}/\mathbb{P}_{j,t}\right)^{\psi}}{\sum_{j'} B_{j',q}\left(W_{j',q,t}/\mathbb{P}_{j',t}\right)^{\psi}}\overline{L}_q. \tag{36}$$

The parameter ψ is the elasticity of employment to real wages. Workers are hand-to-mouth as in the baseline model, so real wages are equivalent to consumption at every instant.

Innovation

Researchers choose both a region i in which to live and a skill level q for which to create varieties. Lab space is distinguished by skill level, so decreasing returns in research labor operate within skill. Knowledge accumulation takes the same form

$$\dot{K}_{i,q,t} = Z_{i,t} N_{i,q,t}^{\lambda} S_{i,q,t}^{\gamma}. \tag{37}$$

Researchers remain hand-to-mouth and move to equalize utility. Utility depends on wages, as in the baseline model, as well as local goods prices $\mathbb{P}_{i,t}$ and region-specific amenities $\mathbb{Q}_i$. Amenities are a reduced-form way of capturing differences in local housing costs. For example, a low $\mathbb{Q}_{\text{San Jose}}$ stands in for expensive housing which, all else equal, deters researchers from working there.[18] In equilibrium, researchers' utility $R_{i,t,q}\frac{\mathbb{Q}_i}{\mathbb{P}_{i,t}}$ assumes a common value.

Diffusion

In the baseline model, researchers draw once-and-for-all diffusion shocks and hold perpetual patents. I now consider a more realistic process of diffusion and imitation.

At constant hazard rate d_{ij}^{K}, an idea from i diffuses to researchers in j, while at constant hazard rate d_{ij}^{X} it becomes available to producers in j. Along the BGP, the shares of i's knowledge available in j to researchers and firms, respectively, are

$$\delta_{ij}^{K} = \frac{d_{ij}^{K}}{d_{ij}^{K} + g^{K}}, \qquad \delta_{ij}^{X} = \frac{d_{ij}^{X}}{d_{ij}^{X} + g^{K}}. \tag{38}$$

[18] Amenities are isomorphic to housing under homothetic preferences and a perfectly elastic housing supply. Income differences between researchers are small, as are consumption and population changes under my counterfactual policy, so this assumption is fairly innocuous.

Any given variety eventually arrives everywhere, although it may take a long time if d is low.

Concurrent with and independent of the diffusion processes, a variety is exogenously imitated at constant hazard rate ν. The imitation shock for a particular variety is common to all regions, and a variety can be imitated before it arrives everywhere. Imitated varieties are supplied at marginal cost by a competitive fringe. As I discuss in the calibration section later, imitation gives the model the realistic feature that the depreciation rate for R&D investment exceeds the risk-free interest rate.

Profits and wages

The set of varieties $M_{j,q,t}$ is the union of two subsets: patented varieties, supplied at markup $\varepsilon/(\varepsilon-1)$ over marginal cost, and imitated varieties, supplied at marginal cost. On the BGP, the share of varieties which are patented is denoted ζ and is equal to $g^K/\left(\nu+g^K\right)$. The effective measure of varieties is $\mathbb{M}_{j,q,t}=\mu M_{j,q,t}$ for a constant $\mu\leq 1$ which depends on model parameters (see Appendix B.8).

The revenue share of patented varieties is denoted η and given by

$$\eta=\frac{\left(\frac{\varepsilon}{\varepsilon-1}\right)^{1-\varepsilon}\zeta}{(1-\zeta)+\left(\frac{\varepsilon}{\varepsilon-1}\right)^{1-\varepsilon}\zeta}. \tag{39}$$

Each patented variety earns flow profits equal to a share ε of flow revenue, while imitated varieties earn no profits. Revenues in a region and skill level are $P_{j,q,t}X_{j,q,t}$, where $X_{j,q,t}\equiv A_{j,q,t}L_{j,q,t}\mathbb{M}_{j,q,t}^{\frac{1}{\varepsilon-1}}$ and the price $P_{j,q,t}$ is consistent with the demand implied by (35). Flow profits are

$$\Pi_{j,q,t}=\frac{\eta}{\zeta\varepsilon}\frac{P_{j,q,t}X_{j,q,t}}{M_{j,q,t}} \tag{40}$$

while wages capture the remaining share of value-added,

$$W_{j,q,t}=\left(1-\frac{\eta}{\varepsilon}\right)\frac{P_{j,q,t}X_{j,q,t}}{L_{j,q,t}}. \tag{41}$$

The variety value (16) is modified to account for diffusion and imitation, and on the BGP takes the form

$$V_{i,t}=\frac{1}{\rho+\nu+g^K}\sum_j\frac{d_{ij}^X}{\rho+\nu+g^K+d_{ij}^X}\Pi_{j,t}.$$

5.2 Equilibrium

Goods market clearing requires that region i's output must meet demand for its good. The market clearing condition is

$$P_{i,t}X_{i,t} = \sum_j \xi_{ij,t}E_{j,t}. \qquad \text{(Market clearing)}$$

where $E_{j,t}$ is consumption expenditure in region j.

The quantitative model features endogenous trade imbalances because factors in one region (researchers and capitalists) have claims on output created in other regions. This means that consumption expenditure is in general not equal to local output. Trade imbalances with costly trade introduce a new motive for spatial redistribution, as the planner is inclined to reallocate researchers across space to manipulate the terms of trade. I judge this channel to be outside the scope of R&D policy, and my model is not intended to study more general spatial reallocation policies. So, I choose to shut down the terms-of-trade reallocation motive by enforcing trade balance. Specifically, I assume that a central authority imposes lump-sum transfers between regional capitalists such that expenditure $E_{j,t}$ is equal to output $P_{j,t}X_{j,t}$. Then, the market clearing condition takes the form

$$P_{i,t}X_{i,t} = \sum_j \xi_{ij,t}P_{j,t}X_{j,t}. \qquad \text{(Market clearing with balanced trade)}$$

Reallocating researchers has no direct effect on aggregate real consumption because transfers offset to ensure market clearing.[19] The only effect is to change real output in different regions.

Nesting the baseline model

To recover the model of Section 3, take $\varphi \to \infty$, $\varkappa_{ij} = 1$, $\psi = 0$, $Q_i = 1$, $\nu = 0$ and return to one worker type.

[19]Local expenditure is equal to labor income (recall that workers and researchers are hand-to-mouth) plus capitalists' expenditure. Capitalists have log utility and so always consume a constant share of their net worth. Lump-sum transfers scale capitalists' net worth up or down without affecting the Euler equation. If a region is research-intensive, its researchers may claim so much income that even deducting capitalists' entire net worth would not restore trade balance. I can handle this case by imposing a global uniform flat labor income tax. This would not affect the equilibrium conditions in any way, and so I do not pursue this extension.

5.3 Research policy in the quantitative model

The optimal research allocation from Proposition 2 extends to the quantitative model. I assume the planner chooses to maximize the present value of log aggregate real consumption, which I define as $\log \sum_j \frac{P_j X_j}{\mathbb{P}_j}$.

Proposition 3 (Place-based policy, quantitative model)**.** *On the BGP, the optimal research allocation is*

$$\frac{S^*_{i,q}}{\bar{S}} = K^*_{i,q,t} \left\{ \varpi \frac{\rho/g^K (1-\lambda)}{1+\rho/g^K} \sum_j X^*_{j,t} \sum_{j'} \frac{\partial \xi^{*\frac{1}{1-\varphi}}_{jj}}{\partial X_{j'}} \left(\frac{X^*_{j',q,t}}{X^*_{j',t}} \right)^{-\frac{1}{\sigma}} \frac{\delta^X_{ij'} X^*_{j',q,t}}{M^*_{j',q,t}} \right.$$
$$+ \varpi \frac{\rho/g^K (1-\lambda)}{1+\rho/g^K} \sum_j \xi^{*\frac{1}{1-\varphi}}_{jj,t} \left(\frac{X^*_{j,q,t}}{X^*_{j,t}} \right)^{-\frac{1}{\sigma}} \frac{\delta^X_{ij} X^*_{j,q,t}}{M^*_{j,q,t}}$$
$$\left. + \frac{\lambda}{1+\rho/g^K} \sum_j \frac{\delta^K_{ij}}{N^*_{j,q,t}} \frac{S^*_{j,q}}{\bar{S}} \right\} \tag{42}$$

where ϖ is a constant ensuring the research labor market clears.

Proposition 3 has two features absent from the simpler Proposition 2. First, the planner accounts for consumption and spillovers within heterogeneous skill types q. Second, the planner considers for trade costs when choosing where to increase consumption and growth. The effects of trade costs are summarized in $\frac{\partial \xi^{*\frac{1}{1-\varphi}}_{jj}}{\partial X_{j'}}$ which, as in Arkolakis, Costinot, and Rodríguez-Clare (2012), governs changes in a region's domestic consumption following output changes in all other regions.

The quantitative model has two sources of misallocation beyond spatial barriers to knowledge flows. First, cost-of-living and amenity differences for researchers mean that free mobility equalizes $R_{i,t,q} \mathbb{Q}_i / \mathbb{P}_{i,t}$, rather than $R_{i,t,q}$. From the demand for research labor, equalizing $R_{i,t,q}$ is tantamount to equalizing the flow of new varieties per researcher $g^K K_{i,q,t} / S_{i,q,t}$, which was the planner's optimality principle in the baseline. Intuitively, if San Jose is an expensive place to live, then there are too few researchers and the marginal product of research labor is too high. Second, and somewhat more subtle, trade costs and spatial sorting make the right-hand side of (42) differ across q. Innovation directed toward a given skill level can have an asymmetric effect on output, hence prices, in different regions based on that skill group's share in regional production.

5.4 Calibration

To quantify the model, I treat the 2019 economy as a balanced growth path equilibrium. The model requires data on (1) employment and wages by city and skill level and (2) patents by city. The labor market and patents data are the same as I used in Section 2. I classify skilled workers as those with a four-year college degree. I feed into the model the 300 largest CBSAs in 2019, which cover 90% of urban employment.

There are no standard public data on city-level research employment. Instead, I use the model to impute the distribution of research employment consistent with free mobility. Here I briefly sketch how, given parameters, I invert the model to recover the unobserved fundamentals—worker productivities $\boldsymbol{A}_\ell$,$\boldsymbol{A}_h$, worker amenities $\boldsymbol{B}_\ell$,$\boldsymbol{B}_h$, and research productivities $\boldsymbol{Z}$—which rationalize the data. Suppose momentarily that one knew the share of patents in each city which were targeted to each of the two skill levels. Production employment and wages yield profits, hence the research wage, for every region and skill level. Free mobility then dictates the supply of researchers implied by that wage, while the knowledge accumulation process is informative of the relative demand for researchers across skill levels within each city. Finally, the relative demand for researchers within each city pins down the share of patents targeted to each skill level. The assumption that researchers' productivities $Z_{i,t}$ and amenities Q_i do not differ by skill is necessary for identification. Otherwise, differences in research fundamentals within a city can generate arbitrary patterns of supply and demand for research labor.

Researcher amenities are not identified from available data, so I assume they are equal to the inverse of the city's regional price parity (BEA 2019). Most of the variation in regional price parities is from housing costs, and so to a first approximation they can be thought of as housing price indices. Modelling amenities this way is very simple, but gives the realistic implication that researchers care about real wages.

The model's key parameters are the spatial barriers to knowledge flows, governed by the arrival rates d_{ij}^K and d_{ij}^X. For other parameters I look to prior literature and use macro moments.

Inferring spatial barriers

I require the model to match the regressions from Section 2 on patent sales and patent citations. Sales inform d_{ij}^X and citations inform d_{ij}^K. Consider sales first. I model the mean of the arrival process as a function of two parameters,

$$d_{ij}^{X}=\begin{cases}1 & i=j\\ \overline{d}^{X}\text{distance}_{ij}^{-\vartheta^{X}} & i\neq j\end{cases}. \tag{43}$$

The choice to set $d_{ii}^{X}=1$ is arbitrary. The regressions from Section 2 are informative about the relative arrival rates between cities, but not the absolute arrival rates. $d_{ii}^{X}=1$ means that, on average, technologies are available locally the year after being invented.

The parameter ϑ^{K} regulates how quickly the arrival rate declines with geographic distance and $\overline{d}^{K}$ pins down the scale of the arrival rate relative to the home region. Consider a variety which is created in region i and licensed to a firm in region j, and recall that the arrival rate is an exponentially distributed random variable. From the properties of the exponential distribution, the probability that a variety from i arrives first in j is the probability that its wait time is the shortest,

$$\Pr\left(\text{first arrive in } j|\text{invented in } i\right)=\frac{d_{ij}^{X}}{\sum_{j'}d_{ij'}^{X}}. \tag{44}$$

I take logs of (44) and add origin and destination fixed effects, which yields a log-linear estimating equation exactly identical to (4) from Section 2. The coefficient on "same city" is the negative log of $\overline{d}^{X}$ (negative because $\overline{d}^{X}$ is for all cities besides the origin), while the coefficient on log distance is ϑ^{K}.

I perform the same exercise with the patent citations regression to recover the arrival rates for the knowledge spillover. My patent citations gravity regression counted only first citations, so as with a sale I can interpret a citation as an initial arrival event.

My preferred specifications, columns (4) of Table 2 and Table 3, give me the four paramters $\overline{d}^{X},\overline{d}^{K},\vartheta^{X},\vartheta^{K}$. With these parameters I compute the $J\times J$ matrices of arrival rates $\boldsymbol{d}^{K}$ and $\boldsymbol{d}^{X}$. To interpret the parameters, first note that $\delta_{ii}^{K}=\delta_{ii}^{X}=0.96$, meaning that local researchers and producers have nearly complete access to local knowledge. This was essentially by construction because the average local arrival lag is only one year. The average external selling barrier δ_{ij}^{X} is 0.55 and the average external learning barrier δ_{ij}^{K} is 0.88. Learning barriers are low, with a typical arrival lag to researchers of just two years. Arrival lags are much longer for selling, with a typical lag of 11 years.[20] Recent evidence

[20]The median arrival lag—that is, the time after which half of varieties have arrived—is 11 years. The mean arrival lag is longer, about 20 years.

on technology diffusion from Bloom et al. (2021) is consistent with long lags. That paper identifies a set of key technologies and traces their diffusion across regions and firms in the US. Bloom et al. (2021) estimate a 16-year gap in average adoption intensity between origin locations and other locations. Other qualitatively consistent evidence comes from the 2018 Annual Business Survey, which included questions on adoption of advanced technologies like robotics, natural language processing, machine vision, and machine learning (Zolas et al. 2021). Adoption rates of new technologies are low, particularly for small and young firms and even when controlling for industry.[21]

Externally calibrated parameters

The trade elasticity $1 - \varphi$ is set to -8. To calibrate the trade costs, I normalize $\varkappa_{ii} = 1$ and set $\varkappa_{ij} = \overline{\varkappa} \cdot \text{distance}_{ij}^{-\kappa}$ for all $j \neq i$. κ is set to match a trade elasticity of -1 given φ, and $\overline{\varkappa}$ is picked to match the average home expenditure share across areas from the 2017 Commodity Flow Survey (CFS), 35%.[22] I follow Acemoglu et al. (2018) and set the elasticity of research output with respect to research labor $\gamma = 0.5$, consistent with empirical evidence on the returns to private R&D.[23] I set the regional labor supply elasticity ψ equal to 2, a consensus value in the literature. The elasticity of substitution between worker types is 1.5, consistent with Acemoglu and Autor 2011. I set the imitation rate ν equal to 0.15, which is a conventional estimate of the depreciation rate of R&D investment (Li and Hall 2020). The discount rate ρ is equal to 0.02.

Macro calibration

I finally set the remaining two parameters: the elasticity of substitution between varieties, ε, and the intertemporal spillover elasticity, λ. With stable labor productivity, then the growth rate of output per worker on the BGP, recall, is equal to the growth rate of the knowledge stock divided by $\varepsilon - 1$. Real output per worker grew at a rate of 1.7% while the number of US-authored patents, is my proxy for knowledge, grew at an annual rate of 4.4% between 1990 and 2019. This implies $\varepsilon = 3.6$, which is consistent with standard estimates of the elasticity of substitution between intermediate inputs.

There is no consensus on the intertemporal spillover across regions. The BGP growth

[21]For example, just 18% of firms in motor vehicle parts manufacturing (NAICS 3363) reported adopting robots, even though auto manufacturing is the most robot-intensive sector.

[22]CFS areas are groups of counties slightly more extensive than my metropolitan areas.

[23]This value is also consistent with a cross-region research labor supply elasticity of 2 under the alternative assumption of heterogeneous research abilities (see Appendix B.1), which matches evidence from Moretti and Wilson (2017).

Table 4: Model parameters

Description	Parameter	Value
Intertemporal spillover	λ	0.54
E.o.S. between varieties	ε	3.6
Returns to scale in research	γ	0.5
E.o.S. between skill groups	σ	1.5
Workers' migration elasticity	ψ	2
Trade elasticity	$1-\varphi$	-8
Elasticity of trade flows wrt distance	$\kappa(1-\varphi)$	-1
Discount rate	ρ	0.02
Obsolescence rate	χ	0.15
Learning barrier (mean)	δ^K	0.88
Learning barrier (s.d.)	δ^K	0.02
Selling barrier (mean)	δ^X	0.55
Selling barrier (s.d.)	δ^X	0.09

rate of knowledge is $g^K = g^Z / (1-\lambda)$, but g^Z encompasses both exogenous increases in research productivity and unmodelled increases in research inputs. I attribute half of the increase in knowledge to changes in research inputs and half to changes in productivity. With research employment growing at 2% annually and $\gamma = 0.5$, this decomposition delivers $\lambda = 0.54$. My calibrated λ is similar toBuera and Oberfield (2020) and nearly identical to the estimate of Arkolakis, Lee, and Peters (2020), who measure the response of patenting to regional variation in immigration flows. Peters (2021) estimates $\lambda = 0.7$ using post-Second World War population movements as a natural experiment.[24]

Table 4 reports the model parameters.

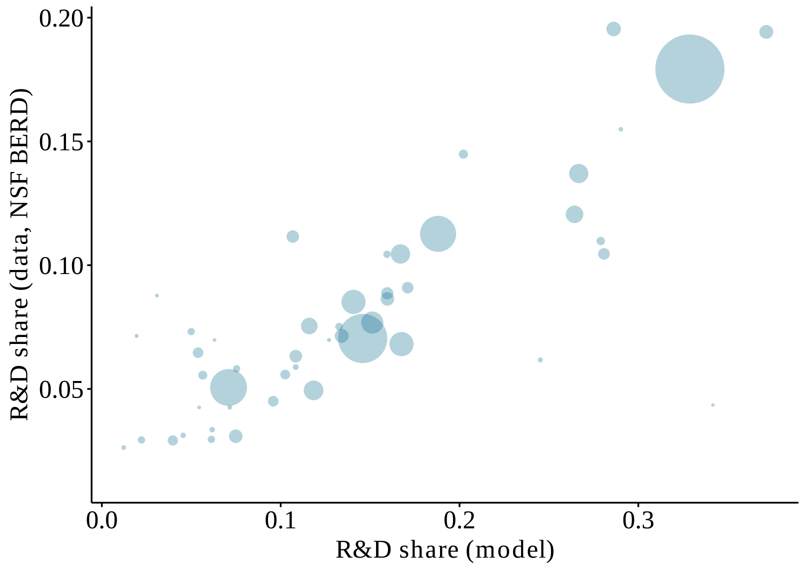
Benchmarking R&D employment

I check the model-implied research employment in two ways. First, I aggregate the model-implied research employment share for 2019 to the state level and compare with NSF Business Enterprise Research and Development (BERD) survey data from 2018 (National Science Foundation 2018). The correlation of state-level R&D employment shares between data and model is 0.91. Second, I construct a proxy for R&D employment using occupational codes in the ACS. I classify occupations as R&D-related based on the list of

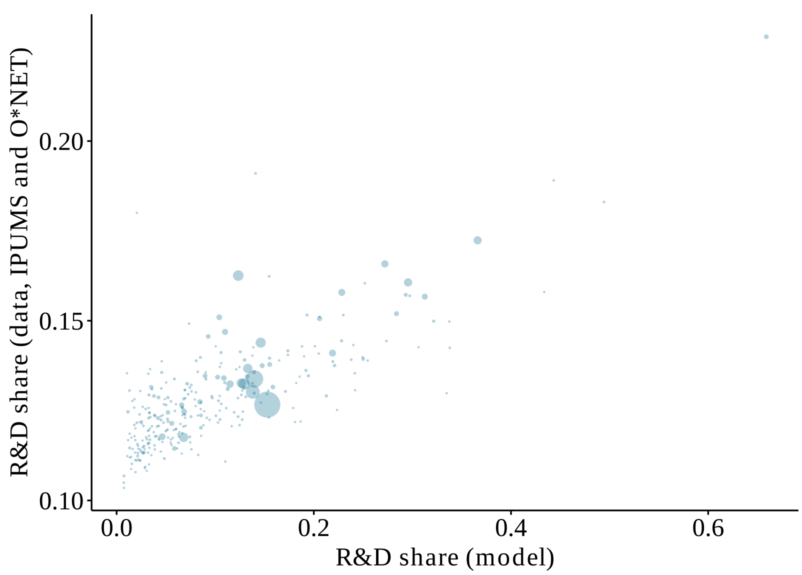
[24]These values are higher than those given by Bloom et al. (2020), who finds $\lambda < 0$ with time-series and industry-level data. In a spatial model, a negative λ would predict that patent rates decline with the local knowledge pool, which is not consistent with my evidence from Section 2.

Figure 3: R&D Employment shares, data vs. model

(a) State-level, NSF data



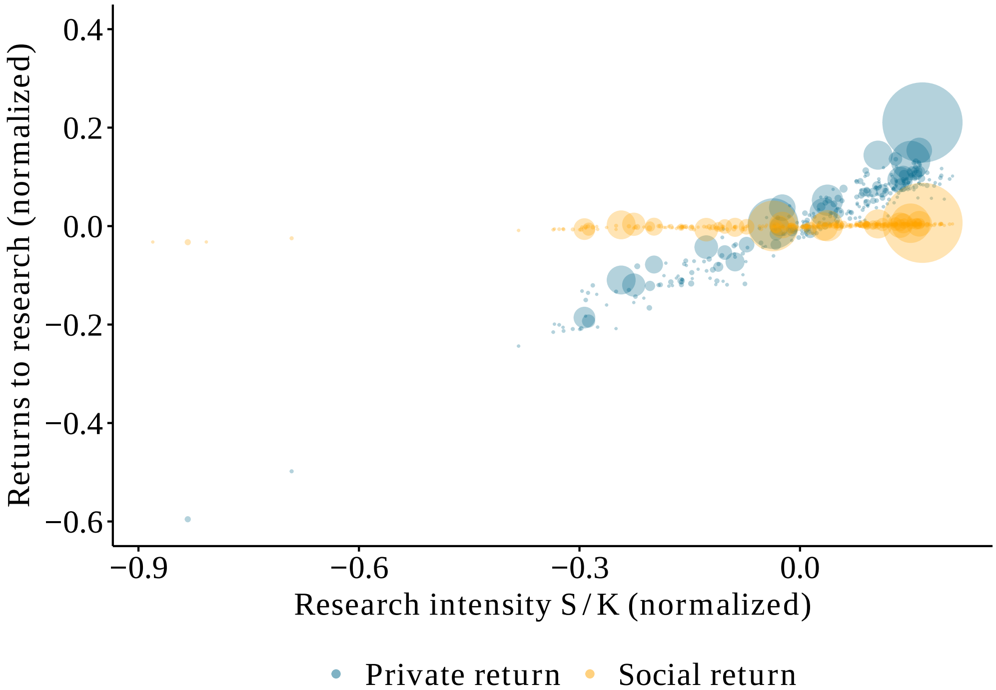
(b) City-level, ACS data



full occupational titles in O*NET, including words like "research" and "design."[25] The correlation of city-level R&D employment shares between data and model is 0.83.

[25] Appendix Table A11 gives the list of occupations.

Figure 4: Research intensity and returns to research



6 Place-based R&D policy

The scope for place-based R&D policy depends on the gap between the private returns to research and the social returns to research. Figure 4 plots the returns to unskilled-complementing research in the baseline equilibrium, where private returns are the log values V_i and the social returns are the log spillover $\sum_j \delta_{ij}^K S_j/N_j$, both normalized to have mean zero. On the horizontal axis I plot the research intensity, which I define as S_i/K_i and which is effectively the planner's choice variable (see the planner's allocation (30)). The size of each marker is proportional to employment. Research intensity is increasing in private returns. Since learning barriers are low and knowledge spillovers vary little across space, research intensity is flat with respect to social returns.

I implement the optimal research allocation with the region- and skill-specific subsidy given in Proposition 3, scaled so that the budget balances. Recall that the optimal policy reallocates researchers to locations with relatively high social returns. Because the social returns are fairly flat across space, the policy ends decongests places with high private returns as shown in Figure 5.

Figure 6 displays a map of the optimal subsidy. Subsidies are high on the West Coast and lowest in the central parts of the country. The West is geographically remote from

Figure 5: Subsidy policy and research intensity

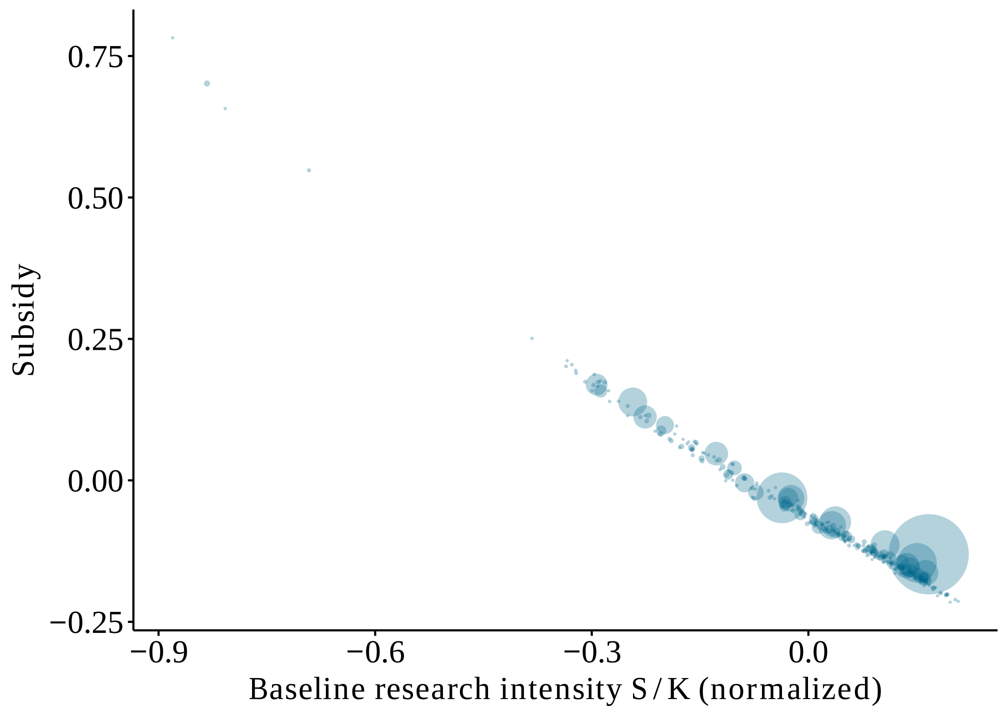


Table 5: Effect of optimal research subsidy

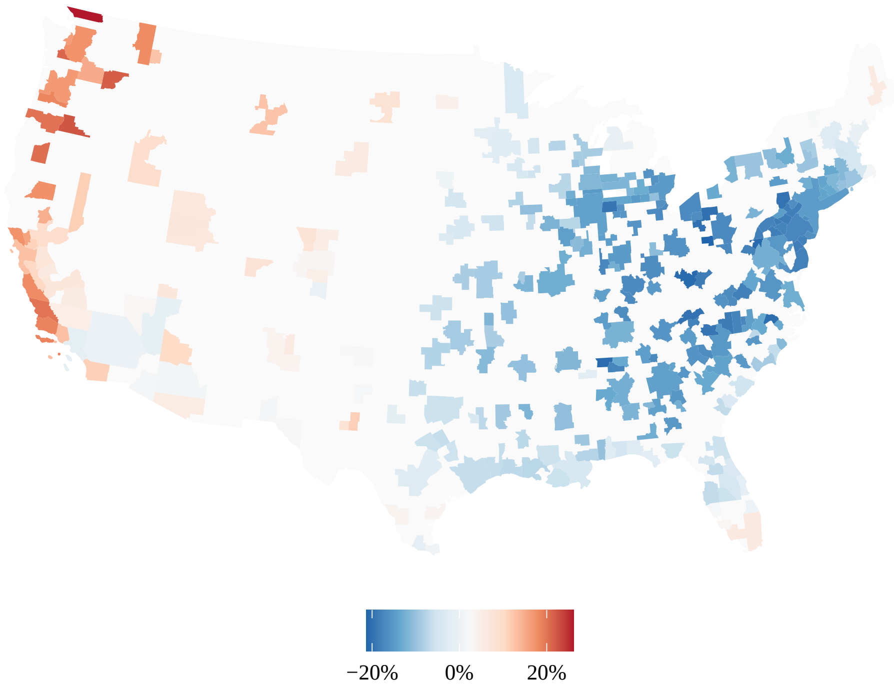
Outcome	Value, relative to baseline
Aggregate consumption	1.008
Aggregate knowledge stock	1.028
Share of scientists re-allocated	11%
Std. dev. log wages, cities	1.002
Std. dev. log patents per worker, cities	1.019
College wage premium	1.000

population centers—about 25% of the US population lives west of the Rocky Mountains—and so the model infers a low private return to research there. I emphasize that this is consistent with the high levels of innovation observed in West Coast research hubs. These hubs create many patents with few researchers, and so the marginal product of research is too high.

The policy increases aggregate consumption by about 0.8% in the long run, as reported in the top panel of Table 5. Growth is driven by a 2.8% increase in the stock of knowledge. About 11% of R&D workers change locations, measured using the formula $\frac{1}{2}\frac{\sum_q \sum_j |S_{j,q} - S^*_{j,q}|}{\bar{S}}$.

The bottom panel of Table 5 explores the distributional consequences of the policy. Wages

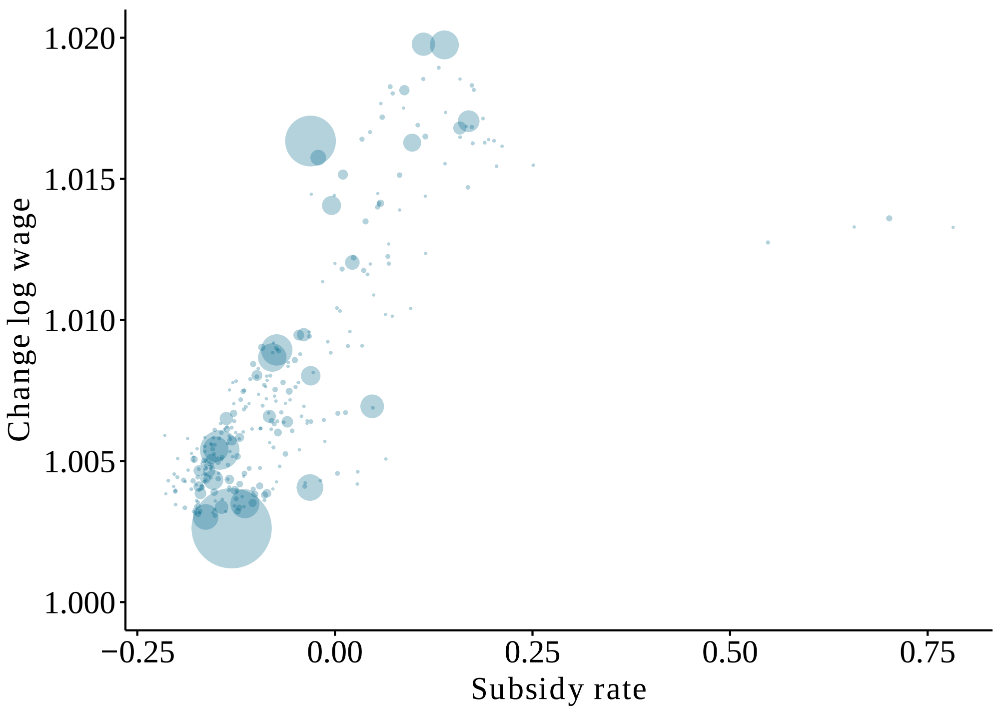
Figure 6: Optimal research subsidy



increase everywhere, but the wage gains are higher in subsidized regions (Figure 7). Spatial wage inequality and especially innovation inequality rise slightly, and the college wage premium does not change. This is somewhat surprising to the extent that small and remote regions, which also have low wages, might be expected to gain researchers. However, private returns are mostly about geography rather than local market size. San Jose, San Francisco, and Seattle all receive subsidies because they are relatively remote from the major US employment centers, and these cities have high wages. The five cities receiving the lowest subsidies (that is, the highest taxes) are in Appalachia and the rural South, because those locations are both centrally located and feature low costs of living. Total patent output is low there, but so is marginal patent output. Remote Denver gets a subsidy while Boston, near large employment centers in the Northeastern US, gets a tax.

West Coast regions also have high costs of living and so there are too few researchers there relative to their marginal product, all else equal. To isolate the effect of market access from amenities, I solve a version of the model in which I allow trade costs and researcher amenities, but impose perfect knowledge diffusion. Reallocation yields a 0.15% gain in

Figure 7: Wage changes and subsidy



aggregate consumption in that environment, less than one-fifth of the 0.8% implied by the full model, and the optimal subsidies are much smaller. I conclude that imperfect knowledge diffusion is a more important source of misallocation in R&D than is cost of living.

6.1 Real-world policy: CHIPS and Science Act

The CHIPS and Science Act of 2022 earmarks $10 billion dollars over five years to establish 20 "regional technology and innovation hubs." The objective is to "create jobs, spur regional economic development, and position communities throughout the country to lead in high-growth, high-wage sectors" (The White House 2022). Reducing aggregate misallocation is not a stated goal of the program, but could in principle be an incidental consequence of redistribution.

I implement a version of the CHIPS policy in my model. The Commerce Department has yet to announce the selection criteria, and so I mimic the policy in the following way. Among cities with a below-median level of patents per worker in 2019, I choose those with the highest share of college graduates in the labor force. I then adjust the site list to comply with the CHIPS Act's geographic diversity requirement.[26] The list of CHIPS-eligible sites is in the first column of Table 6.

CHIPS spending in my model amounts to a 51% research subsidy in the 20 target regions, financed by a 1.5% research tax in all other regions.[27] Research employment grows considerably in subsidized regions, with 82% gains on average, offset by losses on the order of 1.5% to 2% in the other regions.

The CHIPS policy is ineffective. Aggregate real consumption falls by about 0.15% in the long run. Spatial wage inequality decreases very slightly, while spatial inequality in patents per worker goes down by about 1%. Perhaps surprisingly, the policy reduces production workers' wages even in targeted cities. Although patenting goes up by 40% in the average targeted city, patenting falls by about 1% elsewhere. On net, higher knowledge creation in CHIPS cities is more than offset by lower knowledge diffusion from non-CHIPS cities.

The policy fails because the eligibility rule—low patent output and high college shares—is too blunt. As displayed in Table 6, only five of the 20 regions receive positive research subsidies in the optimal policy. These five regions—Anchorage, Honolulu, Olympia, Bismarck, and Sioux Falls—are geographically remote, and so unprofitable places to innovate. The other 15 regions enjoy at least average market access, as indicated by their profits in column 3. They innovate little, but not inefficiently little.

In Appendix C I repeat the exercise with a second interpretation of the CHIPS Act. I use the index constructed by Gruber and Johnson (2019), who argue for regional innovation policy spearheaded in cities with high educational attainment and low cost of living. That policy turns out to be similarly ineffective, for the same reasons as CHIPS: the heuristic policy rule does not discriminate on the margins of inefficiency identified by the model. My findings illustrate the value of a theoretical and quantitative framework for policy design.

[26] The Act requires at least three sites in each of the six Economic Development Administration regions (Economic Development Administration 2022), at least one-third of the sites to have populations under 250,000, and at least one site in a rural state.

[27] Over five years, \$10 billion is \$2 billion per year, about 0.3% of annual US spending on R&D. I assume a uniform subsidy/tax rate within region across skills.

Table 6: CHIPS Act vs optimal policy in twenty regional tech hubs

City	Optimal subsidy	Log profit (relative)	Wage change under CHIPS (%)
Urban Honolulu, HI	0.729	-0.612	-0.161
Anchorage, AK	0.566	-0.509	-0.167
Olympia-Tumwater, WA	0.212	-0.223	-0.181
Bismarck, ND	0.085	-0.132	-0.159
Sioux Falls, SD	0.006	-0.055	-0.149
Traverse City, MI	-0.006	0.013	-0.149
Daphne-Fairhope-Foley, AL	-0.022	-0.023	-0.142
Omaha-Council Bluffs, NE-IA	-0.038	-0.022	-0.142
Charleston-North Charleston, SC	-0.053	0.017	-0.135
Baton Rouge, LA	-0.057	-0.029	-0.142
Wilmington, NC	-0.065	0.020	-0.138
Oklahoma City, OK	-0.076	-0.021	-0.143
Green Bay, WI	-0.086	0.013	-0.148
Little Rock-North Little Rock-Conway, AR	-0.101	0.008	-0.140
Auburn-Opelika, AL	-0.113	0.024	-0.139
Nashville-Davidson–Murfreesboro–Franklin, TN	-0.119	0.068	-0.126
Buffalo-Cheektowaga-Niagara Falls, NY	-0.121	0.068	-0.141
Springfield, IL	-0.127	0.054	-0.144
Richmond, VA	-0.156	0.102	-0.133
California-Lexington Park, MD	-0.160	0.119	-0.143

CHIPS cities are those with the highest college-educated employment shares among cities with below-median patents per worker. Second column refers to the optimal subsidy and third column to the log private value V of research, relative to the average city (both averaged across skill groups). Fourth column is wage change from baseline to policy.

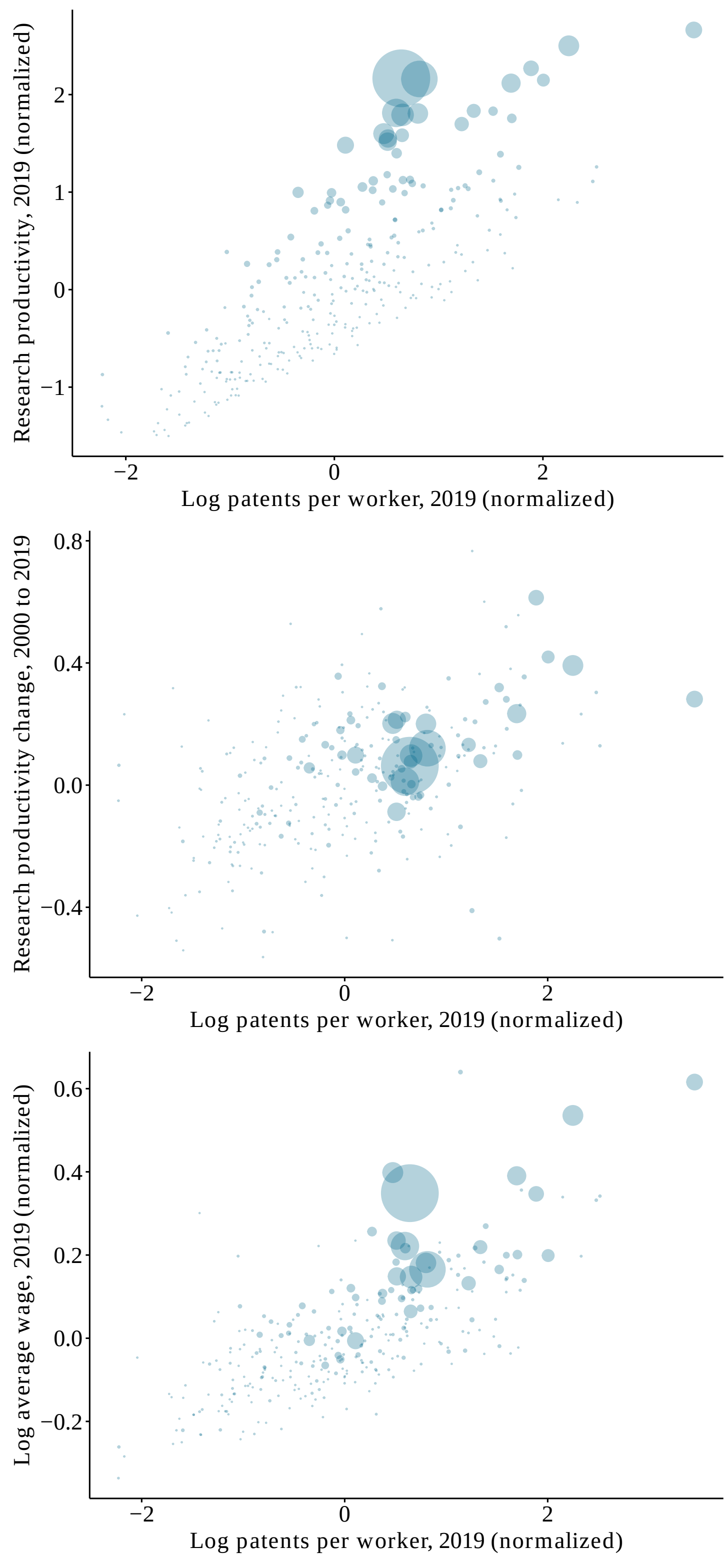
6.2 Why did patenting become more concentrated?

I started the paper by showing that the distribution of patenting became more spatially concentrated over time. The standard deviation across cities of log patents per worker rose by one-sixth between 2000 and 2019. I now conclude by unpacking the distributional effects of rising patent concentration through the lens of my model.

I first solve the model for the 2000 economy. I then consider a counterfactual exercise in which all fundamentals except research productivity $Z_{i,t}$—all workers' productivities, amenities, and the aggregate college employment share—evolved as they did in the data from 2000 to 2019. I hold $Z_{i,t}$ at its 2000 level and solve the new BGP equilibrium.

Research productivity is an important determinant of research output (Figure 8, top panel), and reverting $Z_{i,t}$ back to its 2000 level undoes 80% of the rise in patenting concentration. Regions which experienced falling research productivity had lower patents per worker in 2019 (Figure 8, middle panel). These regions tend to have lower average wages (Figure 8, bottom panel), and so their falling research productivity contributes modestly to rising spatial inequality. The standard deviation of log average wages across cities increases by 6% less in the fixed Z counterfactual than it did in the data. The lion's share of rising inequality came from shocks to fundamentals exogenous to the variety channel in my model, like changes in worker productivity.

Figure 8: Effect of shocks to city-level research productivity



7 Conclusion

When knowledge spills over to future inventors, the social return to innovation exceeds the private return. Spatial barriers to knowledge diffusion introduce a region-specific wedge between social and private returns, motivating region-specific R&D subsidies. In the data, selling patents is more constrained by distance than learning about patents. In consequence, the optimal R&D policy reallocates researchers to regions where research is unprofitable but spillovers are adequate. These regions tend to be remote or have high costs of living. Policy can increase consumption by 0.8% in the long run without devoting more resources to R&D.

References

Acemoglu, Daron. 1998. "Why do new technologies complement skills? Directed technical change and wage inequality." *The Quarterly Journal of Economics* 113 (4): 1055–1089.

———. 2002. "Directed technical change." *The Review of Economic Studies* 69 (4): 781–809.

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. 2018. "Innovation, reallocation, and growth." *American Economic Review* 108 (11): 3450–91.

Acemoglu, Daron, and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of Labor Economics,* 4:1043–1171. Elsevier.

Acemoglu, Daron, and Fabrizio Zilibotti. 2001. "Productivity differences." *The Quarterly Journal of Economics* 116 (2): 563–606.

Akcigit, Ufuk, Murat Alp Celik, and Jeremy Greenwood. 2016. "Buy, keep, or sell: Economic growth and the market for ideas." *Econometrica* 84 (3): 943–984.

Akcigit, Ufuk, Douglas Hanley, and Nicolas Serrano-Velarde. 2021. "Back to basics: Basic research spillovers, innovation policy, and growth." *The Review of Economic Studies* 88 (1): 1–43.

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

Andrews, Michael J, and Alexander Whalley. 2022. "150 years of the geography of innovation." *Regional Science and Urban Economics* 94:103627.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New trade models, same old gains?" *American Economic Review* 102 (1): 94–130.

Arkolakis, Costas, Sun Kyoung Lee, and Michael Peters. 2020. "European immigrants and the United States' rise to the technological frontier." In *2019 Meeting Papers,* vol. 1420.

Armington, Paul S. 1969. "A theory of demand for products distinguished by place of production." *International Monetary Fund Staff Papers* 16 (1): 159–178.

Arora, Ashish, Wesley M Cohen, Honggi Lee, and Divya Sebastian. 2022. "Invention value, inventive capability and the large firm advantage." Technical report. National Bureau of Economic Research.

Arora, Ashish, and Andrea Fosfuri. 2003. "Licensing the market for technology." *Journal of Economic Behavior & Organization* 52 (2): 277–295.

Arque-Castells, Pere, and Daniel F Spulber. 2022. "Measuring the private and social returns to R&D: Unintended spillovers versus technology markets." *Journal of Political Economy* 130 (7): 1860–1918.

Atkeson, Andrew, and Ariel Burstein. 2019. "Aggregate implications of innovation policy." *Journal of Political Economy* 127 (6): 2625–2683.

Atkeson, Andrew, Ariel T Burstein, and Manolis Chatzikonstantinou. 2019. "Transitional dynamics in aggregate models of innovative investment." *Annual Review of Economics* 11:273–301.

Atkin, David, M Keith Chen, and Anton Popov. 2022. "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley." Technical report. National Bureau of Economic Research.

Atkinson, Robert, Mark Muro, and Jacob Whiton. 2019. "The Case for Growth Centers." *Brookings Institution: Washington, DC, USA.*

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018. "Social connectedness: Measurement, determinants, and effects." *Journal of Economic Perspectives* 32 (3): 259–80.

Basu, Susanto, and David N Weil. 1998. "Appropriate technology and growth." *The Quarterly Journal of Economics* 113 (4): 1025–1054.

BEA. 2019. "Metropolitan Regional Price Parities." Dataset. https://www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area.

Berkes, Enrico, and Ruben Gaetani. 2021. "The geography of unconventional innovation." *The Economic Journal* 131 (636): 1466–1514.

Berkes, Enrico, Ruben Gaetani, and Martí Mestieri. 2021. "Cities and Technological Waves." Working paper.

Bloom, Nicholas, Tarek Alexander Hassan, Aakash Kalyani, Josh Lerner, and Ahmed Tahoun. 2021. "The diffusion of disruptive technologies." Working paper. National Bureau of Economic Research.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb. 2020. "Are ideas getting harder to find?" *American Economic Review* 110 (4): 1104–44.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *The Review of Economic Studies* 89 (1): 181–213.

Breschi, Stefano, and Francesco Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of economic geography* 9 (4): 439–468.

Buera, Francisco J, and Ezra Oberfield. 2020. "The global diffusion of ideas." *Econometrica* 88 (1): 83–114.

Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan. 2019. "Migrants, ancestors, and foreign investments." *The Review of Economic Studies* 86 (4): 1448–1486.

Buzard, Kristy, Gerald A Carlino, Robert M Hunt, Jake K Carr, and Tony E Smith. 2017. "The agglomeration of American R&D labs." *Journal of Urban Economics* 101:14–26.

———. 2020. "Localized knowledge spillovers: Evidence from the spatial clustering of R&D labs and patent citations." *Regional Science and Urban Economics* 81:103490.

Cai, Charles, Lorenzo Caliendo, Fernando Parro, and Wei Xiang. 2022. "Mechanics of Spatial Growth." Technical report.

Caselli, Francesco, and Wilbur John Coleman II. 2006. "The world technology frontier." *American Economic Review* 96 (3): 499–522.

Cer, Daniel, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, et al. 2018. "Universal Sentence Encoder." Technical report. https://arxiv.org/abs/1803.11175.

Comin, Diego, and Bart Hobijn. 2010. "An exploration of technology diffusion." *American Economic Review* 100 (5): 2031–59.

Comin, Diego A, Mikhail Dmitriev, and Esteban Rossi-Hansberg. 2012. "The spatial diffusion of technology." Technical report. National Bureau of Economic Research.

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1–45.

Crunchbase. 2021. "Funding Rounds." Dataset. data.crunchbase.com/docs.

Davis, Morris, and Jesse Gregory. 2022. "Place-Based Redistribution in Location Choice Models." Working paper.

Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg. 2018. "The geography of development." *Journal of Political Economy* 126 (3): 903–983.

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

Eckert, Fabian, Teresa C. Fort, Peter K. Schott, and Natalie J. Yang. 2021. "Imputing Missing Values in the US Census Bureau's County Business Patterns." Technical report. National Bureau of Economic Research.

Economic Development Administration. 2022. *EDA REGIONAL OFFICE STRUCTURE.* https://eda.gov/about/.

Fajgelbaum, Pablo D, and Cecile Gaubert. 2020. "Optimal spatial policies, geography, and sorting." *The Quarterly Journal of Economics* 135 (2): 959–1036.

Figueiredo, Octávio, Paulo Guimarães, and Douglas Woodward. 2015. "Industry localization, distance decay, and knowledge spillovers: Following the patent paper trail." *Journal of Urban Economics* 89:21–31.

Figueroa, Nicolás, and Carlos J Serrano. 2019. "Patent trading flows of small and large firms." *Research Policy* 48 (7): 1601–1616.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2014. "Information technology and the distribution of inventive activity." In *The changing frontier: Rethinking science and innovation policy,* 169–196. University of Chicago Press.

Fulford, Scott L, Ivan Petkov, and Fabio Schiantarelli. 2020. "Does it matter where you came from? Ancestry composition and economic performance of US counties, 1850–2010." *Journal of Economic Growth* 25 (3): 341–380.

Glaeser, Edward L, and Naomi Hausman. 2020. "The spatial mismatch between innovation and joblessness." *Innovation Policy and the Economy* 20 (1): 233–299.

Goldschlag, Nathan, Travis J Lybbert, and Nikolas J Zolas. 2020. "Tracking the technological composition of industries with algorithmic patent concordances." *Economics of Innovation and New Technology* 29 (6): 582–602.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik instruments: What, when, why, and how." *American Economic Review* 110 (8): 2586–2624.

Graham, Stuart J. H., Alan C. Marco, and Amanda F. Myers. 2018. "Patent transactions in the marketplace: Lessons from the USPTO Patent Assignment Dataset." *Journal of Economics & Management Strategy* 27 (3): 343–371. doi:https://doi.org/10.1111/jems.12262.

Grossman, Gene M, and Elhanan Helpman. 2018. "Growth, trade, and inequality." *Econometrica* 86 (1): 37–83.

Gruber, Jonathan, and Simon Johnson. 2019. *Jump-starting America: How breakthrough science can revive economic growth and the American dream.* Hachette UK.

Hausman, Naomi. 2022. "University innovation and local economic growth." *The Review of Economics and Statistics* 104 (4): 718–735.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *Quarterly Journal of Economics* 108 (3): 577–598.

Jones, Charles I. 1995. "R & D-based models of economic growth." *Journal of Political Economy* 103 (4): 759–784.

Keller, Wolfgang. 2002. "Geographic localization of international technology diffusion." *American Economic Review* 92 (1): 120–142.

Keller, Wolfgang, and Stephen Ross Yeaple. 2013. "The gravity of knowledge." *American Economic Review* 103 (4): 1414–44.

Kerr, William R. 2008. "Ethnic scientific communities and international technology diffusion." *The Review of Economics and Statistics* 90 (3): 518–537.

Kerr, William R, and Scott Duke Kominers. 2015. "Agglomerative forces and cluster shapes." *Review of Economics and Statistics* 97 (4): 877–899.

Kline, Patrick, and Enrico Moretti. 2014. "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics* 129 (1): 275–331.

Kogan, Leonid, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller. 2021. "Technology-skill complementarity and labor displacement: Evidence from linking two centuries of patents with occupations." Working paper. National Bureau of Economic Research.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. 2017. "Technological innovation, resource allocation, and growth." *The Quarterly Journal of Economics* 132 (2): 665–712.

König, Michael, Kjetil Storesletten, Zheng Song, and Fabrizio Zilibotti. 2022. "From imitation to innovation: Where is all that Chinese R&D going?" *Econometrica* 90 (4): 1615–1654.

Kwon, Hyuk-Soo, Jihong Lee, Sokbae Lee, and Ryungha Oh. 2022. "Knowledge spillovers and patent citations: Trends in geographic localization, 1976–2015." *Economics of Innovation and New Technology* 31 (3): 123–147.

Lentz, Rasmus, and Dale T Mortensen. 2016. "Optimal growth through product innovation." *Review of Economic Dynamics* 19:4–19.

Li, Wendy CY, and Bronwyn H Hall. 2020. "Depreciation of business R&D capital." *Review of Income and Wealth* 66 (1): 161–180.

Li, Zhenlong, Xiao Huang, Tao Hu, Huan Ning, Xinyue Ye, Binghu Huang, and Xiaoming Li. 2021. "ODT FLOW: Extracting, analyzing, and sharing multi-source multi-scale human mobility." *Plos One* 16 (8): e0255259.

Liu, Ernest, and Song Ma. 2021. "Innovation Networks and Innovation Policy." Working paper. National Bureau of Economic Research.

Lychagin, Sergey, Joris Pinkse, Margaret E Slade, and John Van Reenen. 2016. "Spillovers in space: does geography matter?" *The Journal of Industrial Economics* 64 (2): 295–335.

Manuelli, Rodolfo E, and Ananth Seshadri. 2014. "Frictionless technology diffusion: The case of tractors." *American Economic Review* 104 (4): 1368–91.

Moretti, Enrico. 2021. "The effect of high-tech clusters on the productivity of top inventors." *American Economic Review* 111 (10): 3328–75.

Moretti, Enrico, and Daniel J Wilson. 2014. "State incentives for innovation, star scientists and jobs: Evidence from biotech." *Journal of Urban Economics* 79:20–38.

———. 2017. "The effect of state taxes on the geographical location of top earners: Evidence from star scientists." *American Economic Review* 107 (7): 1858–1903.

Moscona, Jacob, and Karthik Sastry. 2022. "Inappropriate Technology: Evidence from Global Agriculture." *Available at SSRN 3886019.*

Müller, Ulrich K, and Mark W Watson. 2021. "Spatial correlation robust inference." *Manuscript, Princeton University.*

National Science Foundation. 2018. "Business Enterprise Research and Development Survey." Dataset. National Center for Science and Engineering Statistics. https://ncses.nsf.gov/pubs/nsf21312.

Parente, Stephen L, and Edward C Prescott. 1994. "Barriers to technology adoption and development." *Journal of Political Economy* 102 (2): 298–321.

Peri, Giovanni. 2005. "Determinants of knowledge flows and their effect on innovation." *Review of Economics and Statistics* 87 (2): 308–322.

Peters, Michael. 2020. "Heterogeneous markups, growth, and endogenous misallocation." *Econometrica* 88 (5): 2037–2073.

———. 2021. "Market size and spatial growth-evidence from germany's post-war population expulsions." Technical report. National Bureau of Economic Research.

Prato, Marta. 2022. "The Global Race for Talent: Brain Drain, Knowledge Transfer, and Economic Growth." Technical report.

Redding, Stephen J, and Esteban Rossi-Hansberg. 2017. "Quantitative spatial economics." *Annual Review of Economics* 9:21–58.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of Political Economy* 98 (5, Part 2): S71–S102.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. 2022. "IPUMS USA." Dataset. Institute for Social Research and Data Innovation.

Saxenian, AnnaLee. 1998. *Silicon Valley's New Immigrant Entrepreneurs.* San Francisco: Public Policy Institute of California.

Schweiger, Helena, Alexander Stepanov, and Paolo Zacchia. 2022. "The long-run effects of R&D place-based policies: evidence from Russian science cities." *American Economic Journal: Economic Policy* 14 (3): 322–51.

Serrano, Carlos J. 2010. "The dynamics of the transfer and renewal of patents." *The RAND Journal of Economics* 41 (4): 686–708.

Service, Internal Revenue. 2022. "SOI Tax Stats: Migration Data." Dataset. https://www.irs.gov/statistics/soi-tax-stats-migration-data.

Souza, Gustavo de. 2022. "The Labor Market Consequences of Appropriate Technology." Technical report. Working Paper.

The White House. 2022. "Fact Sheet: CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China." https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/.

USPTO. 2019. "PatentsView." Dataset.

Walsh, Conor. 2019. "Firm creation and local growth." *Available at SSRN 3496782.*

Zolas, Nikolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David N Beede, Cathy Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz. 2021. "Advanced technologies adoption and use by US firms: Evidence from the annual business survey." Technical report. National Bureau of Economic Research.

A Empirical Appendix

A.1 Patent concentration

I predict patents per worker according to formula (1). To allocate patents to industries, I use the crosswalk provided by Goldschlag, Lybbert, and Zolas (2020), which uses natural language processing techniques to map patent technology classes to industries. I use employment data at the four-digit NAICS level using the County Business Patterns imputation provided by Eckert et al. (2021).

A.2 Bilateral distance measures

Data on migration are tabulated from the IRS Statistics on Income file, 2011-2019 average (Service 2022). Data on the number of trips from origin to destination is from 2019 and is provided by Li et al. (2021), who tabulate Safegraph. Data on Facebook friends is from Facebook's social connectedness index (SCI) (Bailey et al. 2018).

A.3 R&D occupations

I classify as R&D-related those occupations with alternative O*NET titles including any of the following terms: "R&D","research scientist","researcher","research technician","product designer","product design engineer","product development","product developer", and "design engineer." The occupations are listed in Table A11. In addition to the listed occupations, this procedure returned actuaries, lawyers, and registered nurses; I deemed these not R&D-intensive and so chose to remove them.

A.4 Figures and Tables

Table A1: Patent assignments gravity regression, additional specifications

	PPML Log sales			
	(1)	(2)	(3)	(4)
Same city	-0.125***	0.215	0.353***	0.286***
	(0.038)	(0.238)	(0.086)	(0.073)
Industry distance	-3.88**	-2.70	-2.58	-2.02
	(1.84)	(2.32)	(2.11)	(2.45)
Log predicted citations	1.79***	1.86***	1.54***	1.48***
	(0.379)	(0.351)	(0.323)	(0.318)
Log(1+#trips)	0.465***			0.273***
	(0.027)			(0.052)
Log(1+#migrants)		0.376***		-0.002
		(0.027)		(0.025)
Log Facebook friends			0.698***	0.481***
			(0.017)	(0.068)
Log distance				0.211***
				(0.063)
Origin FE	✓	✓	✓	✓
Destination FE	✓	✓	✓	✓
N	779,972	779,972	779,972	779,972
Pseudo R^2	0.92	0.92	0.92	0.92

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the origin state and destination state levels. Models estimated by psuedo-Poisson Maximum Likelihood. Distance is scaled such that the minimum observed log distance is zero and log own-distance is replaced by zero. Industry distance is Euclidean distance of four-digit NAICS level employment shares. Predicted citations based on correlation of patent technology classes.

Table A2: Patent licenses gravity regression

	PPML Log licenses			
	(1)	(2)	(3)	(4)
Same city	3.36***	2.36***	1.91***	1.70***
	(0.350)	(0.276)	(0.662)	(0.642)
Same state		1.11***		
		(0.236)		
Log distance			-0.306***	-0.282***
			(0.108)	(0.107)
Log predicted citations				2.56***
				(0.691)
Industry distance				-1.20
				(2.09)
Origin FE	✓	✓	✓	✓
Destination FE	✓	✓	✓	✓
N	129,490	129,490	129,490	129,490
Pseudo R^2	0.93	0.93	0.93	0.93

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Licensed patents only, sales excluded. Standard errors clustered at the origin state and destination state levels. Models estimated by psuedo-Poisson Maximum Likelihood. Distance is scaled such that the minimum observed log distance is zero and log own-distance is replaced by zero. Industry distance is Euclidean distance of four-digit NAICS level employment shares. Predicted citations based on correlation of patent technology classes.

Table A3: Patent citations gravity regression, additional specifications

	PPML Log citations			
	(1)	(2)	(3)	(4)
Same city	-0.065	0.275**	0.093	0.157**
	(0.065)	(0.123)	(0.076)	(0.079)
Industry distance	-3.67***	-3.26***	-2.84***	-2.29*
	(0.930)	(1.11)	(1.07)	(1.26)
Log predicted citations	2.03***	2.09***	1.94***	1.88***
	(0.150)	(0.155)	(0.150)	(0.151)
Log(1+#trips)	0.197***			0.113***
	(0.005)			(0.023)
Log(1+#migrants)		0.136***		-0.003
		(0.007)		(0.008)
Log Facebook friends			0.313***	0.284***
			(0.004)	(0.028)
Log distance				0.155***
				(0.011)
Citing patent city FE	✓	✓	✓	✓
Cited patent city FE	✓	✓	✓	✓
N	839,972	839,972	839,972	839,972
Pseudo R^2	0.95	0.95	0.95	0.95

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the origin state and destination state levels. Models estimated by psuedo-Poisson Maximum Likelihood. Distance is scaled such that the minimum observed log distance is zero and log own-distance is replaced by zero. Industry distance is Euclidean distance of four-digit NAICS level employment shares. Predicted citations based on correlation of patent technology classes.

Table A4: Distance and bilateral city ties

	OLS Log(1+#migrants)	PPML Log migrants	OLS Log(1+#trips)	PPML Log trips	OLS Log Facebook friends
	(1)	(2)	(3)	(4)	(5)
Log distance	-0.565***	-1.49***	-1.87***	-1.80***	-1.29***
	(0.017)	(0.075)	(0.059)	(0.056)	(0.058)
Same city	9.79***	3.79***	3.11***	1.35***	2.49***
	(0.005)	(0.014)	(0.113)	(0.009)	(0.093)
Origin FE	✓	✓	✓	✓	✓
Destination FE	✓	✓	✓	✓	✓
N	840,889	840,889	840,889	840,889	840,889
Pseudo R^2	0.18	0.99	0.41	0.99	0.45

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the origin state and destination state levels. Distance is scaled such that the minimum observed log distance is zero and log own-distance is replaced by zero.

Table A5: Bartik instrument, first stage

	Δ Neighbors' log patents$_{[t-10,t-1]}$
	(1)
	OLS
Δ Neighbors' predicted log patents$_{[t-20,t-11]}$	1.03^{***}
	(0.036)
Year FE	✓
Observations	1,834
R^2	0.29

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Bartik instrument interacts country f ancestry share in 1900 with lagged growth in country f patenting.

Table A6: US cities' historical ancestry and patent citations

	Log citations
	(1)
	PPML
Country ancestry share$_{1900}$	0.342^{***}
	(0.126)
City FE	✓
Foreign country FE	✓
Observations	35,680

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Citations to foreign $\times$ authored patents granted by the USPTO.

Table A7: Local patent spillovers

	Δ Log patents$_t$	Δ Log wage$_t$	Δ Log real wage$_t$
	(1)	(2)	(3)
		2SLS	
Δ Neighbors' log patents$_{[t-10,t-1]}$	1.92^{***}	0.204^{***}	0.485^{***}
	(0.650)	(0.067)	(0.061)
Δ Neighbors' log employment	-1.50	0.305^{***}	
	(0.984)	(0.075)	
Year FE	✓	✓	✓
Observations	1,834	1,834	1,834
R^2	0.07	0.10	-0.06

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Bartik instrument interacts country f ancestry share in 1900 with lagged growth in country f patenting. Log real wage is log $\frac{wage}{price^{1/3}}$, where price is the city *imes* year fixed effect from a hedonic regression of rents on dwelling characteristics.

Table A8: Patent growth, different distance elasticities

	Δ Log patents$_t$	Δ Log wage$_t$	Δ Log patents$_t$	Δ Log wage$_t$	Δ Log patents$_t$	Δ Log wage$_t$
	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS					
Δ Neighbors' log patents$_{[t-10,t-1]}$	22.9***	5.24***	4.23***	0.828***	0.951***	0.126***
	(7.88)	(0.698)	(1.41)	(0.126)	(0.290)	(0.029)
Distance parameter	-0.1	-0.1	-0.5	-0.5	-2	-2
Year FE	✓	✓	✓	✓	✓	✓
Observations	1,834	1,834	1,834	1,834	1,834	1,834
R^2	0.05	0.02	0.05	0.01	0.04	-0.08

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Bartik instrument interacts patent growth in foreign countries with city historical ancestry. Baseline elasticity is -1.

Table A9: Spatially correlated standard errors

Type	Standard error
Clustered	0.593
Conley (100km cutoff)	0.546
Conley (500km cutoff)	0.303
Mueller-Watson	0.384

Table A10: Local patent spillovers, instrumental variable (European ancestry only)

	Δ Log patents$_t$	Δ Log wage$_t$
	(1)	(2)
	2SLS	
Δ Neighbors' log patents$_{[t-10,t-1]}$	2.06**	0.349***
	(0.841)	(0.080)
Year FE	✓	✓
Observations	1,834	1,834
R^2	0.05	-0.00
KP F stat	278.59	278.59

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Bartik instrument interacts country f ancestry share in 1900 with lagged growth in country f patenting. Western European countries only (Austria, Belgium, Luxembourg, Switzerland, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Iceland, Italy, Netherlands, Norway, Portugal, and Sweden).

Table A11: List of R&D occupations

Occupation	Code (`occ2010`)
Architectural and engineering managers	300
Operations research analysts	1220
Statisticians	1230
Aerospace engineers	1320
Chemical engineers	1350
Civil engineers	1360
Computer hardware engineers	1400
Electrical and electronics engineers	1410
Materials engineers	1450
Mechanical engineers	1460
Petroleum, mining and geological engineers, including mining safety engineers	1520
Engineers, nec	1530
Agricultural and food scientists	1600
Medical scientists, and life scientists, all other	1650
Astronomers and physicists	1700
Chemists and materials scientists	1720
Environmental scientists and geoscientists	1740
Physical scientists, nec	1760
Social scientists, nec	1840
Agricultural and food science technicians	1900
Biological technicians	1910
Chemical technicians	1920
Clinical laboratory technologists and technicians	3300

Figure A1: Variation across cities in patents per worker and average wages

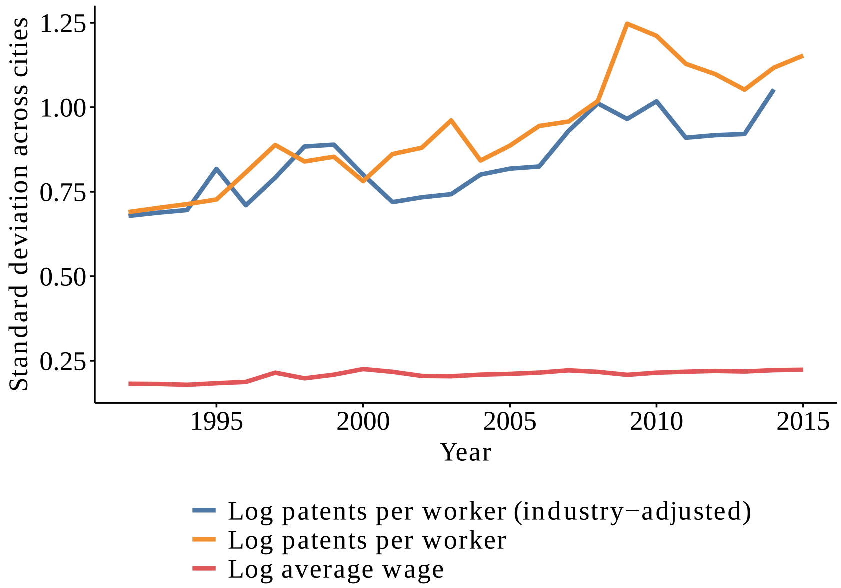
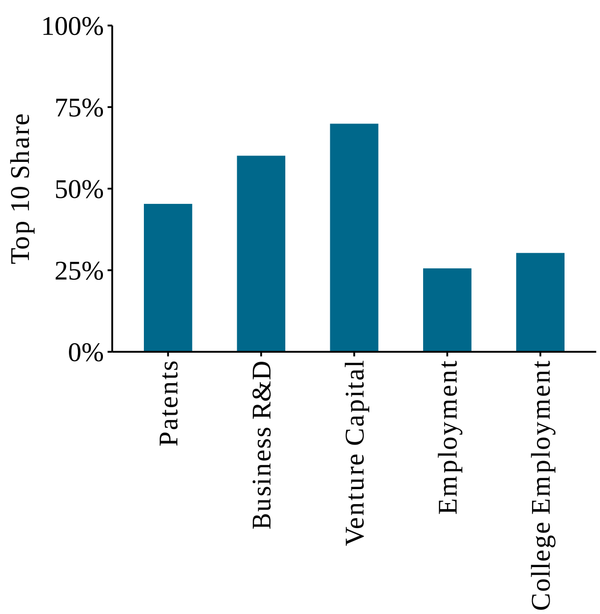


Figure A2: Top 10 cities perform a high share of R&D



"Top 10" are the ten cities producing the most patents from 2000 to 2019: San Jose, San Francisco, New York, Los Angeles, Boston, Seattle, San Diego, Chicago, Minneapolis, and Dallas. *Sources*: US Patent and Trademark Office (patents), Census and American Community Survey (employment and educational attainment), National Science Foundation Business and Enterprise Research Survey (business R&D), and Crunchbase (venture capital).

B Theory Appendix

B.1 Local congestion in research

In the main text I assumed research labs operated under decreasing returns to scale. Here I show two alternative interpretations of local congestion

B.1.1 Heterogeneous research ability

Research labs produce under constant returns to scale but workers have heterogeneous research ability in each region. Each researcher can supply y_i units of research labor to labs in region i. y_i is a Fréchet-distributed random variable with shape $\epsilon > 1$ and scale unity. The CDF is $F\left(y_i\right) = \exp\left(-y_i^{-\epsilon}\right)$. Researchers sort on productivity. If the wage is R_i, the number of researchers in region i is

$$L_i^S = \left(\frac{R_i}{\mathbb{R}}\right)^{\epsilon} \overline{L}^S \tag{45}$$

where I define $\mathbb{R} = \left(\sum_i R_i^{\epsilon}\right)^{1/\epsilon}$. Total labor supplied, S_i, is the number of researchers multiplied by the average quantity of labor supplied per researcher. The average quantity of labor supplied per researcher is

$$\begin{aligned} \overline{y}_i &= \mathbb{E}\left(y_i | R_i y_i \geq R_{i'} y_{i'} \; \forall i \neq i'\right) && (46) \\ &= \Gamma\left(1 - \frac{1}{\epsilon}\right) \times \frac{\mathbb{R}}{R_i} && (47) \\ &= \Gamma\left(1 - \frac{1}{\epsilon}\right) \times \left(\frac{L_i^S}{\overline{L}^S}\right)^{-1/\epsilon} && (48) \end{aligned}$$

where the second line follows the standard Fréchet algebra, the third line uses (45), and $\Gamma\left(\cdot\right)$ is the gamma function. Then,

$$\begin{aligned} S_i &= L_i^S \times \overline{y}_i \\ &\propto \left(L_i^S\right)^{1-1/\epsilon} \end{aligned}$$

with $1 - 1/\epsilon \in (0, 1)$ corresponding to γ in (10).

B.1.2 Scarce local factor

Suppose that variety creation is a Cobb-Douglas function of research labor S and an intermediate input Q, which I call lab space. Production is constant returns to scale with labor share γ. A representative lab creates a flow of varieties equal to

$$S_j^{\gamma} Q_j^{1-\gamma} Z_j N_j^{\lambda}$$

Lab space is supplied competitively with marginal cost given by

$$F_j = Q_j^{\beta} \tag{49}$$

for a parameter $\beta > 0$. The lab sells each variety for price V_j and chooses inputs to maximize profits (which are zero in equilibrium by constant returns to scale). Substituting the marginal cost curve (49) into the demand for lab space yields the equilibrium quantity of lab space,

$$Q_j^* = \left((1-\gamma) Z_j S_j^{\gamma} N_j^{\lambda} V_j\right)^{\frac{1}{1+\beta}} \tag{50}$$

Substituting (50) into the demand curve for researchers then yields, after some algebra,

$$S_j = \widetilde{\gamma} R_j^{-\frac{\beta+\gamma}{\beta(1-\gamma)}} \left(Z_j N_j^{\lambda} V_j\right)^{\frac{1}{\beta}+\frac{\beta+\gamma}{\beta(1-\gamma)}} \tag{51}$$

for a constant $\widetilde{\gamma}$. This is a more general instance of the demand curve (18) in the main text and nests that specification when $\beta \to \infty$. The main text corresponds to $Q_j^* = 1$, which is the only value consistent with positive and finite marginal cost for $\beta \to \infty$.

B.2 Characterization of BGP equilibrium

In this section I derive the growth rate of all variables on the balanced growth path (BGP). The BGP is characterized by a constant researcher allocation S and a constant interest rate ι. First, rewrite the knowledge accumulation equation (10) as

$$g^{K_{i,t}} K_{i,t} = N_{i,t}^{\lambda} Z_{i,t} S_{i,t}^{\gamma}. \tag{52}$$

On the BGP, K must grow at a constant rate g^K. Imposing this and then differentiating (52) gives

$$g^K = \lambda g^N + g^Z. \tag{53}$$

Then, observe that $g^N = g^K$ because N is a linear combination of the $\boldsymbol{K}$. So,

$$g^K = \frac{g^Z}{1-\lambda} \tag{54}$$

where, recall, g^Z is exogenous. Next, $g^M = g^K$ because M is a linear combination of the $\boldsymbol{K}$. Log differentiating the expression for output (7) gives $g^X = \frac{1}{\varepsilon-1} g^M = \frac{1}{\varepsilon-1} g^K$. Doing the same for wages and profits gives $g^W = \frac{1}{\varepsilon-1} g^K = g^X$ and $g^\Pi = \frac{1}{1-\varepsilon} g^K - g^K = g^X - g^K$. The value per variety grows at the same rate as profits, $g^V = g^\Pi$. Then, log differentiating the labor demand curve for researchers (18) and fixing S gives $g^V + \lambda g^N + g^Z = g^R$, which simplifies to $g^\Pi + g^K = g^X = g^R$. Lastly, the rental price for lab space grows at rate $g^F = g^V + g^K = g^X$ because lab space rents command a share $1-\gamma$ of labs' revenue, which is proportional to $V \times K$.

Putting everything together, Z grows exogenous rate g^Z. Knowledge stocks grow at rate $g^K = \frac{1}{\varepsilon-1} g^Z$. Consumption and income—wages, aggregate profits, rents—grow at rate $g^X = g = \frac{1}{\varepsilon-1} g^K$.

Finally, note that the BGP knowledge stocks $K_{i,t}$ are the unique solution to the knowledge accumulation equation

$$g^K K_{i,t} = \left(\sum_j \delta^K_{ji} K_{j,t} \right)^\lambda S_i^\gamma Z_{i,t},$$

which is a non-linear system of J equations in the J unknowns $\boldsymbol{K}(t)$. In particular, the equilibrium $\boldsymbol{K}(t)$ consistent with these equations are independent of the initial knowledge stocks $K_{i,0}$.

B.3 Proof of Lemma 1

Consider a small shock to the research allocation, $d \log S_i$. Totally log differentiate the growth equation (10) to get

$$d \log K_i = \lambda \sum_j \Omega^K_{ji} d \log K_j + \gamma d \log S_i,$$

or, in matrix form,

$$\log \boldsymbol{K} = \lambda \Omega^{K'} d \log \boldsymbol{K} + \gamma d \log \boldsymbol{S}.$$

Rearrange this expression to obtain (26). For the second part of the Lemma, totally log differentiate (9) to obtain

$$d \log W_i = \frac{1}{\varepsilon - 1} \sum_j \Omega^X_{ji} d \log K_j,$$

which is (27).

B.4 Proof of Proposition 1

Proof. I specialize the formula (26) to the case of no learning barriers. The learning absorption matrix is $\Omega^K_{ij} = \frac{K_i}{N}$, where N takes the common value $N = \sum_{i'} K_{i'}$. I use the following lemma:

Lemma 2 (Sherman-Morrison formula)**.** *Let u and v be column vectors in $\mathbb{R}^J$ and let $A = uv'$. Denote the trace of A as $\mathrm{Tr}\,(A)$. If $\mathrm{Tr}\,(A) \neq 1$, then*

$$(I - A)^{-1} = I + \frac{A}{1 - \mathrm{Tr}\,(A)}$$

Apply the Lemma with u a column vector of ones and $v_i = -\lambda K_i / N$. I obtain $\left(I - \lambda \Omega^{K'}\right)^{-1} = I + \frac{\lambda}{1-\lambda} \Omega^{K'}$. Substitute into (26) to obtain, in summation form,

$$d \log K_i = \gamma d \log S_i + \gamma \frac{\lambda}{1 - \lambda} \frac{1}{N} \sum_j K_j d \log S_j. \tag{55}$$

In the case of no selling barriers, $K_j \propto S_j$ by (24). Then $\sum_j K_j d \log S_j \propto \sum_j S_j d \log S_j = 0$, where the equality follows because $\sum_j S_j$ is fixed at $\overline{S}$. Thus, $d \log K_i = \gamma d \log S_i$.

For the second part of the Proposition, no selling barriers imply $d \log X_j$ and $d \log W_j$ are each proportional to $d \log M$, where $M = N = \sum_i K_i$. This derivative takes the form

$$\begin{aligned} d \log M &= \frac{1}{M} \sum_j K_j d \log K_j \\ &= \frac{\gamma}{M} \sum_j K_j d \log S_j \\ &\propto \sum_j S_j d \log S_j \\ &= 0 \end{aligned}$$

□

B.5 Proof of Proposition 2

Proof. Without loss of generality, start at time 0 with some initial knowledge stocks $K_{i,0}$. The problem is

$$\max_{\{K_{i,t}\},\{S_{i,t}\}} \int_0^\infty \exp(-\rho t) \log\left(\sum_j X_{j,t}\right) dt$$

subject to

$$\dot{K}_{i,t} = N_{i,t}^{\lambda} S_{i,t}^{\gamma} Z_{i,t}, \qquad \sum_i S_{i,t} = \overline{S}$$

Set up the current-value Hamiltonian with a Lagrange multiplier on the researcher constraint,

$$\mathcal{H}(\boldsymbol{K}, \boldsymbol{S}, t, \boldsymbol{\mu}, \chi) = \log \sum_j X_{j,t} + \sum_j \mu_{j,t} N_{i,t}^{\lambda} S_{i,t}^{\gamma} Z_{i,t} + \chi_t \left(\overline{S} - \sum_i S_{i,t}\right)$$

The necessary conditions are:

1. $\mathcal{H}_{S_{i,t}} = 0$ for all i, t
2. $\rho\mu_{i,t} - \dot{\mu}_{i,t} = \mathcal{H}_{K_{i,t}}$ for all i, t
3. $\dot{K}_{i,t} = N_{i,t}^{\lambda} S_{i,t}^{\gamma} Z_{i,t}$ and $\overline{S} = \sum_i S_{i,t}$ for all i, t
4. $\lim_{t\to\infty} \exp(-\rho t)\, \mu_{i,t} K_{i,t} = 0$

The first condition reads

$$\begin{aligned} \gamma \mu_{i,t} N_{i,t}^{\lambda} Z_{i,t} S_{i,t}^{\gamma-1} - \chi_t &= 0 \\ \Rightarrow \mu_{i,t} &= \frac{\chi_t}{\gamma N_{i,t}^{\lambda} Z_{i,t} S_{i,t}^{\gamma-1}} \end{aligned}$$

The second condition reads

$$\begin{aligned} \rho\mu_{i,t} - \dot{\mu}_{i,t} = \mathcal{H}_{K_{i,t}} &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\frac{\partial M_{j,t}}{\partial K_{i,t}} + \lambda \sum_j \frac{\partial N_{j,t}}{\partial K_{i,t}} \mu_{j,t} N_{j,t}^{\lambda-1} Z_{j,t} S_{j,t}^{\gamma} \\ &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^{X} + \lambda \sum_j \delta_{ij}^{K} \mu_{j,t} N_{j,t}^{\lambda-1} Z_{j,t} S_{j,t}^{\gamma} \end{aligned}$$

Substitute the first condition into the second to get

$$\begin{aligned}\rho\mu_{i,t} - \dot{\mu}_{i,t} &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^X + \lambda\sum_j \delta_{ij}^K \frac{\chi_t}{\gamma N_{i,t}^{\lambda} Z_{i,t} S_{i,t}^{\gamma-1}} N_{j,t}^{\lambda-1} Z_{j,t} S_{j,t}^{\gamma} \\ &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^X + \frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}\delta_{ij}^K \end{aligned} \tag{56}$$

Next, use the third condition rewrite the first condition in the form

$$\mu_{i,t} = \frac{\chi_t S_{i,t}}{\gamma g^K K_{i,t}}$$

Differentiating this with respect to time implies

$$g^{\mu_{i,t}} = g^{\chi_t} - g^K \tag{57}$$

where g^K is given exogenously on the BGP. I restrict attention to the BGP solution. Using (57) I rewrite (56) as

$$\begin{aligned}\rho\mu_{i,t} - g^{\mu_{i,t}}\mu_{i,t} &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^X + \frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}\delta_{ij}^K \\ \left(\rho + g^K - g^{\chi_t}\right)\mu_{i,t} &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^X + \frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}\delta_{ij}^K \\ \left(\rho + g^K - g^{\chi_t}\right)\frac{\chi_t S_{i,t}}{\gamma g^K K_{i,t}} &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^X + \frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}\delta_{ij}^K \\ \Rightarrow \left(\rho + g^K - g^{\chi_t}\right)\frac{\chi_t S_{i,t}}{\gamma g^K} &= K_{i,t}\left\{\frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^X + \frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}\delta_{ij}^K\right\} \end{aligned} \tag{58}$$

To solve for χ_t, sum both sides over i to obtain

$$\begin{aligned}\left(\rho + g^K - g^{\chi_t}\right)\frac{\chi_t \overline{S}}{\gamma g^K} &= \sum_i K_{i,t}\frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\delta_{ij}^X + \sum_i K_{i,t}\frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}\delta_{ij}^K \\ &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}\sum_i \delta_{ij}^X K_{i,t} + \frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}\sum_i K_{i,t}\delta_{ij}^K \\ &= \frac{1}{\varepsilon-1}\frac{1}{X_t}\sum_j \frac{X_{j,t}}{M_{j,t}}M_{j,t} + \frac{\lambda\chi_t}{\gamma}\sum_j \frac{S_{j,t}}{N_{j,t}}N_{j,t} \\ &= \frac{1}{\varepsilon-1} + \frac{\lambda\chi_t}{\gamma}\overline{S} \end{aligned}$$

Rearranging yields

$$\chi_t = \frac{1}{\overline{S}} \frac{\gamma}{\varepsilon - 1} \frac{1}{\frac{\rho - g^{\chi_t}}{g^K} + 1 - \lambda} \tag{59}$$

Subsitute (59) back in to (58) to obtain

$$\frac{S_{i,t}}{\overline{S}} = \left\{ \frac{\rho/g^K + (1-\lambda) - g^{\chi_t}/g^K}{\rho/g^K + 1 - g^{\chi_t}/g^K} \sum_j \frac{X_{j,t}}{X_t} \frac{\delta_{ij}^X K_{i,t}}{M_{j,t}} + \frac{\lambda}{\rho/g^K + 1 - g^{\chi_t}/g^K} \sum_j \frac{S_{j,t}}{\overline{S}} \frac{\delta_{ij}^K K_{i,t}}{N_{j,t}} \right\}$$

$S_{i,t}$ must be constant on the BGP, as must each of the two sums. In the knife-edge case $\sum_j \frac{X_{j,t}}{X_t} \frac{\delta_{ij}^X K_{i,t}}{M_{j,t}} = \sum_j \frac{S_{j,t}}{\overline{S}} \frac{\delta_{ij}^K K_{i,t}}{N_{j,t}}$, then the right-hand side does not depend on g^{χ_t} and the result goes through as stated. In all other cases, the right-hand side is constant if and only if g^{χ_t} is constant. From (59), g^{χ_t} is constant only if it is zero. I then know $g^{\mu_t} = -g^K$. It is straightforward to verify that the transversality condition is satisfied.

I maintain $\varepsilon \geq 2$, which is sufficient (although not necessary) to guarantee that the maximized Hamiltonian is strictly concave. The given solution is therefore the unique global optimum. □

B.6 Proof of Corollary 1

Proof. Researchers earn a share γ of research revenue, so

$$R_{i,t} S_{i,t} = \gamma V_{i,t} g^K K_{i,t}$$

The net wage is

$$\begin{aligned} (1+\varsigma_i) R_{i,t} &= (1+\varsigma_i) \frac{\gamma V_{i,t} g^K K_{i,t}}{S_{i,t}} \\ &= \overline{R}_t \end{aligned}$$

where the second line is implied by free mobility. Substitute the expression for the value $V_{i,t}$, (23), to obtain

$$\begin{aligned} \overline{R}_t &= (1+\varsigma_i) \gamma g^K \frac{1}{\rho + g_K} \frac{1}{\varepsilon} \frac{K_{i,t}}{S_{i,t}} \sum_j \frac{\delta_{ij}^X X_{j,t}}{M_{j,t}} \\ \Rightarrow 1 + \varsigma_i &\propto \frac{S_{i,t}}{K_{i,t}} \left(\sum_j \frac{\delta_{ij}^X}{M_{j,t}} \frac{X_{j,t}}{X_t} \right)^{-1} \end{aligned}$$

Next, use the optimal $S^*_{i,t}$, (30), to get

$$1 + \varsigma_i \propto \left(\omega \sum_j \frac{\delta^X_{ij}}{M^*_{j,t}} \frac{X^*_{j,t}}{X^*_t} + (1-\omega) \sum_j \frac{\delta^K_{ij}}{N^*_{j,t}} \frac{S^*_{j,t}}{\overline{S}} \right) \left(\sum_j \frac{\delta^X_{ij}}{M^*_{j,t}} \frac{X^*_{j,t}}{X^*_t} \right)^{-1}$$

where I have defined $\omega \equiv \frac{\rho/g^K + (1-\lambda)}{\rho/g^K + 1}$. Multiply through to get

$$\begin{aligned} 1 + \varsigma_i &\propto \omega + (1-\omega) \frac{\sum_j \frac{\delta^K_{ij}}{N^*_{j,t}} \frac{S^*_{j,t}}{\overline{S}}}{\sum_j \frac{\delta^X_{ij}}{M^*_{j,t}} \frac{X^*_{j,t}}{X^*_t}} \\ &\propto 1 + \frac{1-\omega}{\omega} \frac{\sum_j \frac{\delta^K_{ij}}{N^*_{j,t}} \frac{S^*_{j,t}}{\overline{S}}}{\sum_j \frac{\delta^X_{ij}}{M^*_{j,t}} \frac{X^*_{j,t}}{X^*_t}} \\ &= 1 + \frac{\lambda}{\rho/g^K + 1 - \lambda} \frac{\sum_j \frac{\delta^K_{ij}}{N^*_{j,t}} \frac{S^*_{j,t}}{\overline{S}}}{\sum_j \frac{\delta^X_{ij}}{M^*_{j,t}} \frac{X^*_{j,t}}{X^*_t}} \end{aligned}$$

□

B.7 Model Extensions

B.7.1 Trade in varieties

In the baseline model I assumed that the technology to produce a variety was bought and sold, but that the varieties themselves were non-traded. Here I consider a different specification of the model's production side. Rather than license their varieties to other producers, inventors produce their varieties locally and export them. I consider a BGP and suppress time subscripts.

An inventor creates a variety in region i. The variety can be sold to the competitive final good producer (equivalently, directly to consumers) in each region j subject to iceberg trade costs $\varkappa_{ij} \leq 1$: a consumer in j receives only a fraction $\varkappa_{ij}$ of each unit shipped from a supplier in i. The final good producer combines traded varieties in a CES fashion to produce the final good, which is then consumed locally. I do not alter the process of variety creation; in particular, the knowledge spillover process is the same as in the baseline model.

Each producer has a unit labor requirement $1/A_i$ and can hire labor competitively at wage W_i. Employment is fixed at L_i. The producer will choose a constant markup over marginal cost, setting a price $P_i = \frac{\varepsilon}{\varepsilon-1}\frac{W_i}{A_i}$. The variety price in region j is then $P_i/\varkappa_{ij}$ and the consumption price index in region j is

$$\mathbb{P}_j^{1-\varepsilon} = \sum_i M_i \left(P_i/\varkappa_{ij}\right)^{1-\varepsilon}$$

To close the trade module I impose balanced trade through regional transfers, either between immobile production workers or immobile capitalists. Trade balance means that goods consumption in region j, $\mathbb{P}_j\mathbb{X}_j$, is equal to goods production in region j. Goods production is pinned down by the labor market clearing condition; the quantity of output is equal to the labor endowment A_jL_j and each unit of output is sold at price P_j, so production is equal to $P_jA_jL_j$. This effectively pins down the index of real consumption $\mathbb{X}_j$. The equilibrium is a vector of prices $\boldsymbol{P}$ such that goods markets clear.

Let the measure of varieties be M_i. The share of j's spending on varieties from region i is

$$\xi_{ij} = \frac{M_i \left(P_i/\varkappa_{ij}\right)^{1-\varepsilon}}{\mathbb{P}_j^{1-\varepsilon}}$$

which then implies that revenue to region i firms is

$$\begin{aligned} P_iA_iL_i &= \sum_j \xi_{ij}\mathbb{P}_j\mathbb{X}_j \\ &= \sum_j \frac{M_i \left(P_i/\varkappa_{ij}\right)^{1-\varepsilon}}{\mathbb{P}_j^{1-\varepsilon}} P_jA_jL_j \end{aligned}$$

The flow profits for a producer in i are a share ε of local revenue per variety,

$$\begin{aligned} \Pi_i &= \frac{1}{\varepsilon}\frac{P_iA_iL_i}{M_i} \\ &= \frac{1}{\varepsilon}\left(\frac{A_iL_i}{M_i}\right)^{1-\frac{1}{\varepsilon}} \left(\sum_j \frac{\varkappa_{ij}^{\varepsilon-1}}{\mathbb{P}_j^{1-\varepsilon}} P_jA_jL_j\right)^{\frac{1}{\varepsilon}} \end{aligned}$$

Profits are increasing in local market size A_iL_i, and decreasing in the measure of competing local varieties M_i.

In the special case of free trade, profits depend only on local variables,

$$\Pi_i \propto \left(\frac{A_i L_i}{M_i}\right)^{1-\frac{1}{\varepsilon}}$$

Free mobility in research dictates

$$\frac{S_i}{M_i} \propto V_i \propto \Pi_i$$

As in the baseline model, researchers fail to internalize the intertemporal knowledge spillovers they grant to others. The difference here is that the reallocation motive persists even with free trade and perfect knowledge diffusion. Profits per variety are low in small markets, so by free mobility the flow of varieties per marginal researcher is high. The planner values these extra varieties because of learning.

B.7.2 Static agglomeration

Consider static agglomeration in researcher productivity. The research accumulation equation (10) is now

$$\dot{K}_{i,t} = S_{i,t}^{\gamma} N_{i,t}^{\lambda} \mathbb{Z}_{i,t}$$

where I have replaced $Z_{i,t}$ with $\mathbb{Z}_{i,t} = Z_{i,t} S_{i,t}^{\beta}$ for a parameter $\beta > 0$. I assume $\gamma + \beta < 1$ so that local agglomeration is not strong enough to create a black hole in research. The agglomeration is not internalized by research labs.

The decentralized solution goes through as in the baseline model. The planner's problem has the same structure except that the planner accounts for static agglomeration in research. (See Appendix B.5 for the baseline Hamiltonian problem.) In particular, the first Hamiltonian necessary condition is modified to

$$(\gamma + \beta)\, \mu_{i,t} N_{i,t}^{\lambda} Z_{i,t} S_{i,t}^{\gamma+\beta-1} - \chi_t = 0$$

It is straightforward to then follow the same steps as in B.5. The optimality condition turns out to be exactly the same as (30). In particular, in the case with no diffusion barriers, $\delta_{ij}^{X} = \delta_{ij}^{K} = 1$. Then the right-hand side of (30) is constant, and so the optimality condition says $S_i^* = \varpi_t K_{i,t}^*$ at every point in time for an endogenous constant ϖ_t. By comparison, the decentralized allocation with agglomeration satisfies

$$\begin{aligned}\gamma S_{i,t}^{\gamma-1+\beta} N_{i,t}^{\lambda} Z_{i,t} V_t &= R_t \\ \Rightarrow S_{i,t} &= \frac{\gamma g^K K_{i,t} V_t}{R_t}\end{aligned} \tag{60}$$

where V_t does not vary by region because of perfect diffusion and the second line follows from the knowlege accumulation equation. Condition (60) coincides with the planner's optimum so the decentralized allocation is efficient. This extension demonstrates that static local agglomeration is fundamentally different from dynamic regional spillovers.

Moretti (2021) estimates the elasticity of inventor output to research cluster size using inventor microdata and an instrumental variables design. His preferred estimate is $\widehat{\beta} = 0.07$ (see Moretti (2021) Table 3, column 8) and he finds that the constant-elasticity functional form is a good description of the data.[28] Moretti (2021) argues that reallocating researchers increases aggregate patent output even without diffusion barriers, which differs from my conclusion. The reason for the difference is that I assume free mobility in research without compensating differentials, which equalizes the private and social returns to research across space in the frictionless case.

B.7.3 Large research labs

Consider the problem of a large research lab. The lab is large in the sense that its knowledge stock in region i, which I denote by $k_{i,t}$, is non-negligible relative to the aggregate knowledge stock $K_{i,t}$. For the purpose of exposition I suppose the lab faces internal barriers to knowledge flows. In practice these are likely to be smaller than the barriers I estimate at the level of the aggregate economy.

The lab has a given endowment of researchers, $\bar{s}$, and allocates them across regions to maximize the present discounted value of profits. Denote profits $y_t \equiv \sum_j y_{j,t}$ where $y_{j,t} \equiv m_{j,t}\pi_{j,t}$ is total profit in each region. The problem is

$$\max_{\{s_{i,t}\}} \int_0^\infty \exp(-\rho t) \log y_t dt$$

[28] The cluster is defined at the level of the research field (for example, computer science). My model could be interpreted as encompassing a large number of research fields operating Cobb-Douglas production functions over labor and land, so that knowledge accumulation is constant returns to scale at the field level but decreasing returns to scale at the city level. The empirical analysis of Moretti (2021) looks at variation within city across fields.

subject to the law of motion for the lab's knowledge stocks

$$\dot{k}_{i,t} = N_{i,t}^{\lambda} s_{i,t}^{\gamma} Z_{i,t}$$

The lab's stock of varieties available in j is $m_{j,t} = \sum_i \delta_{ij}^X k_{i,t}$, and profits per variety are $\pi_{j,t} = \frac{1}{\varepsilon}\frac{X_{j,t}}{M_{j,t}}$. I assume that the lab cannot cannot oligopolize licensing, so cannot influence markups. However, the lab internalizes the effect of its own decisions on aggregate innovation, profits, and spillovers.

The setup of the problem is as in Appendix B.5 and I restrict attention to the BGP solution. The first two necessary Hamiltonian conditions are

1. $\mu_{i,t} Z_{i,t} \gamma N_{i,t}^{\lambda} s_{i,t}^{\gamma-1} = \chi_t$
2. $\rho\mu_{i,t} - \dot{\mu}_{i,t} = \frac{1}{y_t}\sum_j \frac{\partial\left(m_{j,t}\pi_{j,t}\right)}{\partial k_{i,t}} + \sum_j \mu_{j,t}\frac{\partial \dot{k}_{j,t}}{\partial k_{i,t}}$

Consider the second condition. The first term on the right-hand side is

$$\begin{aligned}
\frac{\partial\left(m_{j,t}\pi_{j,t}\right)}{\partial k_{i,t}} &= \frac{\partial}{\partial k_{i,t}}\left(\frac{1}{\varepsilon}\frac{X_{j,t}}{M_{j,t}} m_{j,t}\right) \\
&= \frac{1}{\varepsilon}\left\{\frac{\frac{\partial X_{j,t}}{\partial k_{i,t}} m_{j,t} + \frac{\partial m_{j,t}}{\partial k_{i,t}} X_{j,t}}{M_{j,t}}\right\} - \frac{1}{\varepsilon}\frac{\frac{\partial M_{j,t}}{\partial k_{i,t}} X_{j,t} m_{j,t}}{M_{j,t}^2} \\
&= \frac{1}{\varepsilon}\delta_{ij}^X\left\{\frac{\frac{1}{\varepsilon-1}\frac{X_{j,t}}{M_{j,t}} m_{j,t} + X_{j,t}}{M_{j,t}}\right\} - \frac{1}{\varepsilon}\frac{\delta_{ij}^X X_{j,t} m_{j,t}}{M_{j,t}^2} \\
&= \frac{1}{\varepsilon}\delta_{ij}^X\left\{\frac{\frac{1}{\varepsilon-1} X_{j,t} m_{j,t} + X_{j,t} M_{j,t} - X_{j,t} m_{j,t}}{M_{j,t}^2}\right\} \\
&= \frac{1}{\varepsilon}\delta_{ij}^X \frac{X_{j,t}}{M_{j,t}}\left\{1 + \frac{2-\varepsilon}{\varepsilon-1}\frac{m_{j,t}}{M_{j,t}}\right\}
\end{aligned}$$

The derivative of the knowledge accumulation equation is

$$\begin{aligned}
\frac{\partial \dot{k}_{j,t}}{\partial k_{i,t}} &= \frac{\partial}{\partial k_{i,t}} N_{j,t}^{\lambda} s_{j,t}^{\gamma} Z_{j,t} \\
&= \lambda N_{j,t}^{\lambda-1} \delta_{ij}^K s_{j,t}^{\gamma} Z_{j,t}
\end{aligned}$$

Putting it together gives

$$\left(\rho - g^{\mu_{i,t}}\right) \mu_{i,t} = \frac{1}{y_t}\frac{1}{\varepsilon}\sum_j \delta^X_{ij}\frac{X_{j,t}}{M_{j,t}}\left\{1 + \frac{2-\varepsilon}{\varepsilon-1}\frac{m_{j,t}}{M_{j,t}}\right\} + \sum_j \mu_{j,t}\lambda N^{\lambda-1}_{j,t}\delta^K_{ij}s^{\gamma}_{j,t}Z_{j,t}$$

The first condition plus the knowledge accumulation equation give

$$\mu_{i,t} = \frac{\chi_t s_{i,t}}{\gamma k_{i,t}}$$

and by the same reasoning as in B.5, $g^{\mu_{i,t}} = -g^K$ and χ_t is constant. So,

$$\left(\rho + g^K\right)\frac{\chi s_{i,t}}{g^K\gamma k_{i,t}} = \frac{1}{y_t}\frac{1}{\varepsilon}\sum_j \delta^X_{ij}\frac{X_{j,t}}{M_{j,t}}\left\{1 + \frac{2-\varepsilon}{\varepsilon-1}\frac{m_{j,t}}{M_{j,t}}\right\} + \frac{\lambda\chi}{\gamma}\sum_j \delta^K_{ij}\frac{s_{j,t}}{N_{j,t}} \tag{61}$$

To focus the exposition on knowledge spillovers, I specialize to the case $\varepsilon = 2$. When $\varepsilon = 2$, profits per variety do not depend on the measure of varieties, and so the lab's behavior does not affect the level of profits per variety. (61) then simplifies to

$$\left(\rho + g^K\right)\frac{\chi s_{i,t}}{g^K\gamma k_{i,t}} = \frac{1}{y_t}\frac{1}{\varepsilon}\sum_j \delta^X_{ij}\frac{X_{j,t}}{M_{j,t}} + \frac{\lambda\chi}{\gamma}\sum_j \delta^K_{ij}\frac{s_{j,t}}{N_{j,t}} \tag{62}$$

Next, define the lab-level usage absorption and learning absorption matrices with entries

$$\omega^X_{ij} = \frac{\delta^X_{ij}k_{i,t}}{m_{j,t}} \qquad \omega^K_{ij} = \frac{\delta^K_{ij}k_{i,t}}{n_{j,t}}$$

which I use to rewrite (62) as

$$\begin{aligned}
\left(\rho + g^K\right)\frac{\chi}{g^K\gamma}s_{i,t} &= \frac{1}{y_t}\frac{1}{\varepsilon}\sum_j \frac{\delta^X_{ij}k_{i,t}}{m_{j,t}}m_{j,t}\frac{X_{j,t}}{M_{j,t}} + \frac{\lambda\chi}{\gamma}\sum_j \frac{\delta^K_{ij}k_{i,t}}{n_{j,t}}n_{j,t}\frac{s_{j,t}}{N_{j,t}} \\
&= \frac{1}{y_t}\frac{1}{\varepsilon}\sum_j \omega^X_{ij}m_{j,t}\frac{X_{j,t}}{M_{j,t}} + \frac{\lambda\chi}{\gamma}\sum_j \omega^K_{ij}n_{j,t}\frac{s_{j,t}}{N_{j,t}} \\
&= \frac{1}{y_t}\sum_j \omega^X_{ij}m_{j,t}\pi_{j,t} + \frac{\lambda\chi}{\gamma}\sum_j \omega^K_{ij}n_{j,t}\frac{s_{j,t}}{N_{j,t}} \\
&= \frac{1}{y_t}\sum_j \omega^X_{ij}y_{j,t} + \frac{\lambda\chi}{\gamma}\sum_j \omega^K_{ij}n_{j,t}\frac{s_{j,t}}{N_{j,t}}
\end{aligned} \tag{63}$$

Next, I sum (63) over i to solve for χ and find

$$\begin{aligned}\left(\rho+g^{K}\right)\frac{\chi\bar{s}}{g^{K}\gamma}&=\frac{1}{y_t}\sum_j y_{j,t}\sum_i \omega_{ij}^{X}+\frac{\lambda\chi}{\gamma}\sum_j\frac{s_{j,t}}{N_{j,t}}\sum_i\delta_{ij}^{K}k_{i,t}\\&=1+\frac{\lambda\chi}{\gamma}\sum_j\frac{s_{j,t}}{N_{j,t}}n_{j,t}\end{aligned}$$

Define the lab's effective spillover share $\phi_i\left(\bar{s}\right)\equiv\sum_j\omega_{ij}^{K}\frac{s_j}{\bar{s}}\frac{n_{j,t}}{N_{j,t}}$, and define $\Phi\left(\boldsymbol{s}\right)=\sum_i\phi_i\left(\boldsymbol{s}\right)$. Rearranging the preceding line delivers

$$\begin{aligned}\frac{\chi\bar{s}}{\gamma}\frac{\rho+g^{K}}{g^{K}}&=1+\frac{\lambda\chi}{\gamma}\bar{s}\Phi\left(\boldsymbol{s}\right)\\\Rightarrow\frac{\gamma}{\chi}&=\bar{s}\left(\frac{\rho+g^{K}}{g^{K}}-\lambda\Phi\left(\boldsymbol{s}\right)\right)\end{aligned}$$

which I substitute back into (63) to get

$$\begin{aligned}\frac{\rho+g^{K}}{g^{K}}s_{i,t}&=\frac{\gamma}{\chi}\sum_j\omega_{ij}^{X}\frac{y_{j,t}}{y_t}+\lambda\sum_j\omega_{ij}^{K}n_{j,t}\frac{s_{j,t}}{N_{j,t}}\\\Rightarrow\frac{s_{i,t}}{\bar{s}}&=\frac{g^{K}}{\rho+g^{K}}\frac{1}{\bar{s}}\frac{\gamma}{\chi}\sum_j\omega_{ij}^{X}\frac{y_{j,t}}{y_t}+\lambda\frac{g^{K}}{\rho+g^{K}}\sum_j\omega_{ij}^{K}\frac{s_{j,t}}{\bar{s}}\frac{n_{j,t}}{N_{j,t}}\\\Rightarrow\frac{s_{i,t}}{\bar{s}}&=\left(1-\lambda\frac{g^{K}}{\rho+g^{K}}\Phi\left(\boldsymbol{s}\right)\right)\sum_j\omega_{ij}^{X}\frac{y_{j,t}}{y_t}+\lambda\frac{g^{K}}{\rho+g^{K}}\phi_i\left(\bar{s}\right)\end{aligned}$$

which characterizes the lab's optimal researcher allocation $s_{i,t}^{*}$. When $n_{j,t}/N_{j,t}\to 1$ such that the lab takes over the research sector, then $\Phi=1$ and the lab behaves just like the social planner, balancing profits and spillovers. In the polar case $n_{j,t}/N_{j,t}\to 0$ such that the lab's knowledge pool becomes small relative to the total knowledge pool, then $\Phi,\phi_i\to 0$. The lab chooses to ignore spillovers, with $\frac{s_{i,t}}{\bar{s}}=\sum_j\omega_{ij}^{X}\frac{y_{j,t}}{y_t}$ based on profits only. For intermediate cases the lab places positive weight on spillovers, but less than the social planner.

B.8 Details of Quantitative Model

Knowledge diffusion shares

Let $N_{ij,t}$ be the stock of knowledge from region i which has diffused to inventors in region j by time t. The stock of knowledge yet to diffuse is $K_{i,t}-N_{ij,t}$, and ideas within that set diffuse at constant hazard d_{ij}^K. On a BGP all stocks grow at BGP rate g^K. Therefore,

$$\begin{aligned}\dot{N}_{ij,t} &= d_{ij}^K \left(K_{i,t} - N_{ij,t}\right) \\ \Rightarrow g^K N_{ij,t} &= d_{ij}^K \left(K_{i,t} - M_{ij,t}\right) \\ \Rightarrow \frac{N_{ij,t}}{K_{i,t}} &= \frac{d_{ij}^K}{d_{ij}^K + g^K},\end{aligned}$$

which is the left-hand side of (38). The right-hand side follows by an entirely symmetric argument for producers, replacing d_{ij}^K with d_{ij}^X.

Effective meausure of varieties

I suppress time, region, and skill-type subscripts. Out of the stock M of varieties, M^R are patent-protected and M^I are imitated. Relative demand is given by $\frac{x^I}{x^R} = \left(\frac{p^I}{p^R}\right)^{-\varepsilon}$, where x^I (x^R) is the quantity demanded for an imitated (patented) variety. Patented varieties are priced at markup $\varepsilon / (\varepsilon - 1)$ while imitated varieties are priced at cost, so $\frac{x^I}{x^R} = \left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon}$. The demand for labor must clear the labor market,

$$\begin{aligned}M^I x^I + M^R x^R &= AL \\ \Rightarrow M^I \left(\frac{\varepsilon - 1}{\varepsilon}\right)^{-\varepsilon} x^R + M^R x^R &= AL \\ \Rightarrow x^R &= \frac{AL}{M^I \left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} + M^R}\end{aligned}$$

and

$$x^I = \left(\frac{\varepsilon - 1}{\varepsilon}\right)^{-\varepsilon} \frac{AL}{M^I \left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} + M^R}$$

This means output is given by

$$
\begin{aligned}
X^{\frac{\varepsilon-1}{\varepsilon}} &= \int_0^{M^I} x^I(m)^{\frac{\varepsilon-1}{\varepsilon}}\,dm + \int_0^{M^R} x^R(m)^{\frac{\varepsilon-1}{\varepsilon}}\,dm \\
&= \int_0^{M^I} \left(\frac{AL\left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon}}{M^I\left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} + M^R} \right)^{\frac{\varepsilon-1}{\varepsilon}} dm + \int_0^{M^R} \left(\frac{AL}{M^I\left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} + M^R} \right)^{\frac{\varepsilon-1}{\varepsilon}} dm \\
&= \left(\frac{AL}{M^I\left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} + M^R} \right)^{\frac{\varepsilon-1}{\varepsilon}} \left\{ M^I\left(\frac{\varepsilon-1}{\varepsilon}\right)^{1-\varepsilon} + M^R \right\} \\
\Rightarrow X &= \frac{AL}{M^I\left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} + M^R} \left(M^I\left(\frac{\varepsilon-1}{\varepsilon}\right)^{1-\varepsilon} + M^R \right)^{\frac{\varepsilon}{\varepsilon-1}} \\
&= AL\mathbb{M}^{\frac{1}{\varepsilon-1}}
\end{aligned}
$$

with

$$
\mathbb{M} = M \underbrace{\left((1-\zeta)\left(\frac{\varepsilon-1}{\varepsilon}\right)^{1-\varepsilon} + \zeta \right)^{\varepsilon} \left((1-\zeta)\left(\frac{\varepsilon-1}{\varepsilon}\right)^{-\varepsilon} + \zeta \right)^{1-\varepsilon}}_{\equiv \mu}
$$

It is straightforward to verify that $\mu \in (0,1]$, attaining the upper bound for $\nu = 0$. Next, revenue is PX is split between wages and profits. Only patented varieties earn profits, while imitators earn no profits and dedicate their full revenues to the wage bill.

The share of the final goods producer's expenditure on patented varieties is

$$
\eta = \frac{M^R\left(p^R\right)^{1-\varepsilon}}{M^R\left(p^R\right)^{1-\varepsilon} + M^I\left(p^I\right)^{1-\varepsilon}}
$$

which delivers (39) after substituting $p^R/p^I = \varepsilon/(\varepsilon-1)$ and $M^R/M^I = \zeta/(1-\zeta)$. Profits to patented varieties are a share $1/\varepsilon$ of revenues ηPX; the measure of patented varieties is ζM, so profit per patented variety is

$$
\Pi = \frac{\eta}{\varepsilon}\frac{PX}{\zeta M}
$$

which is (40). The wage bill consumes the remainder of value-added,

$$
WL = PX - \Pi\zeta M = \left(1 - \frac{\eta}{\varepsilon}\right) PX
$$

Value of a variety

The present discounted value of a variety created in region i at time t takes the form

$$V_{i,t} = \sum_j \int_t^\infty \exp\left(-\iota\left(\tau - t\right)\right) \times \left(1 - \exp\left(-d_{ij}^X\left(\tau - t\right)\right)\right) \times \left(\exp\left(-\chi\left(\tau - t\right)\right)\right) \Pi_{j,\tau} d\tau$$

which reflects that flow profits are discounted at rate ι, the diffusion shock arrives at rate d_{ij}^X, and the variety is imitated at rate χ.

Profits grow at rate g^Π, so

$$V_{i,t} = \sum_j \int_t^\infty \exp\left(-\iota\left(\tau - t\right)\right) \times \left(1 - \exp\left(-d_{ij}^X\left(\tau - t\right)\right)\right) \times \exp\left(-\chi\left(\tau - t\right)\right) \exp\left(g^\Pi\left(\tau - t\right)\right) d\tau$$

The integral evaluates to

$$\begin{aligned} V_{i,t} &= \sum_j \left(\frac{1}{\iota - g^\Pi + \chi} - \frac{1}{\iota - g^\Pi + \chi + d_{ij}^X}\right) \Pi_{j,t} \\ &= \frac{1}{\iota - g^\Pi + \chi} \sum_j \frac{d_{ij}^X}{\iota - g^\Pi + \chi + d_{ij}^X} \Pi_{j,t} \end{aligned}$$

and plugging in the Euler equation gives

$$V_{i,t} = \frac{1}{\rho + g^K + \chi} \sum_j \frac{d_{ij}^X}{\rho + g^K + \chi + d_{ij}^X} \Pi_{j,t}$$

B.9 Proof of Proposition 3

Define aggregate real consumption by $U_t \equiv \sum_j \frac{P_{j,t} X_{j,t}}{\mathbb{P}_{j,t}}$. The planner's problem is

$$\max_{\left\{S_{i,q,t}\right\}} \int_0^\infty \exp\left(-\rho t\right) \log U_t dt$$

subject to the laws of motion for knowledge (37), the production function (35), employment (36), trade balance, and the aggregate researcher constraint $\sum_{i,q} S_{i,q,t} = \overline{S}$.

Proof. The first two necessary Hamiltonian conditions are $\partial\mathcal{H}/\partial S_{i,q,t} = 0$ and $\mu_{i,q,t} - \dot{\mu}_{i,q,t} = \frac{\partial\mathcal{H}}{\partial K_{i,q,t}}$. The third condition is the law of motion for knowledge and the fourth is the transversality condition.

The first condition is unchanged and gives

$$\mu_{i,q,t} = \frac{\chi}{\gamma S_{i,q,t}^{\gamma-1} N_{i,q,t}^{\lambda} Z_i} = \frac{\chi_t S_{i,q,t}}{\gamma g^K K_{i,q,t}}$$

It is convenient to write the real output price as a function of the own-trade share, $P_{j,t}/\mathbb{P}_{j,t} = \xi_{jj}^{\frac{1}{1-\varphi}}$. Then, $\log U_t = \log\left(\sum_j \xi_{jj,t}^{\frac{1}{1-\varphi}} X_{j,t}\right)$ and the second Hamiltonian condition is

$$\mu_{i,q,t} - \dot{\mu}_{i,q,t} = \frac{1}{U_t}\left\{\sum_j X_{j,t}\frac{\partial \xi_{jj,t}^{\frac{1}{1-\varphi}}}{\partial K_{i,q,t}} + \sum_j \xi_{jj,t}^{\frac{1}{1-\varphi}}\frac{\partial X_{j,t}}{\partial K_{i,q,t}}\right\} + \sum_j \mu_{j,q,t}\lambda N_{j,q,t}^{\lambda-1} S_{j,q,t}^{\gamma} Z_{j,t} \tag{64}$$

By identical reasoning to the proof in B.5, an admissible BGP solution requires $g^{\chi_t} = 0$ and $g^{\mu_{i,t}} = -g^K$; otherwise, the right-hand side of (64) is unstable. These growth rates satisfy the transversality condition.

The two derivatives are

$$\frac{\partial X_{j,t}}{\partial K_{i,q,t}} = X_{j,t}^{\frac{1}{\sigma}} X_{j,q,t}^{-\frac{1}{\sigma}} \frac{1}{\varepsilon-1}\frac{X_{j,q,t}}{M_{j,q,t}}\delta_{ij}^X$$

and

$$\begin{aligned}\frac{\partial \xi_{jj,t}^{\frac{1}{1-\varphi}}}{\partial K_{i,q,t}} &= \sum_{j'} \frac{\partial \xi_{jj,t}^{\frac{1}{1-\varphi}}}{\partial X_{j',t}}\frac{\partial X_{j',t}}{\partial K_{i,q,t}} \\ &= \frac{1}{\varepsilon-1}\sum_{j'} \frac{\partial \xi_{jj,t}^{\frac{1}{1-\varphi}}}{\partial X_{j',t}} X_{j',t}^{\frac{1}{\sigma}} X_{j',q,t}^{-\frac{1}{\sigma}} \frac{X_{j',q,t}}{M_{j',q,t}}\delta_{ij'}^X\end{aligned}$$

I evaluate (64) following the same steps as in B.5, substituting out for $\mu_{j,q,t}$ with the first Hamiltonian condition and solving for χ with the aggregate researcher constraint. This delivers (42) in the text. □

Implementation details

To implement the policy, I compute $\frac{\partial \xi_{jj}^{\frac{1}{1-\varphi}}}{\partial X_{j'}}$ using the implicit function theorem. Let the function $F_i(\boldsymbol{X}, \boldsymbol{P}) = P_i X_i - \sum_j \xi_{ij} P_j X_j$ describe the trade balance condition. The relation

$\boldsymbol{F}(\boldsymbol{X}, \boldsymbol{P}) = 0$ implicitly defines a function $\boldsymbol{P} = \boldsymbol{G}(\boldsymbol{X})$. By the implicit function theorem,

$$\left[\frac{\partial \boldsymbol{P}}{\partial \boldsymbol{X}}\right] = \left[\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{X}}\right] = -\left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{P}}\right]^{-1}\left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{X}}\right]$$

Next,

$$\xi_{jj}^{\frac{1}{1-\varphi}} = \frac{P_j}{\mathbb{P}_j}$$

so

$$\frac{\partial \xi_{jj}^{\frac{1}{1-\varphi}}}{\partial X_i} = \frac{\partial P_j}{\partial X_i}\frac{\mathbb{P}_j}{\mathbb{P}_j^2} - \frac{\partial \mathbb{P}_j}{\partial X_i}\frac{P_j}{\mathbb{P}_j^2}$$

where

$$\mathbb{P}_j = \left(\sum_{i'}\left(P_{i'}/\varkappa_{i'j}\right)^{1-\varphi}\right)^{\frac{1}{1-\varphi}}$$

Hence,

$$\begin{aligned}
\frac{\partial \mathbb{P}_j}{\partial X_i} &= \sum_{i'}\frac{\partial \mathbb{P}_j}{\partial P_{i'}}\frac{\partial P_{i'}}{\partial X_i} \\
&= \mathbb{P}_j \mathbb{P}_j^{\varphi-1}\sum_{i'}\frac{1}{P_{i'}}P_{i'}^{1-\varphi}\varkappa_{i'j}^{\varphi-1}\frac{\partial P_{i'}}{\partial X_i} \\
&= \mathbb{P}_j\sum_{i'}\frac{1}{P_{i'}}\xi_{i'j}\frac{\partial P_{i'}}{\partial X_i}
\end{aligned}$$

Putting it together:

$$\begin{aligned}
\left[\frac{\partial \xi_{jj}^{\frac{1}{1-\varphi}}}{\partial X_{j'}}\right]_{ji} &= \frac{\partial \xi_{jj}^{\frac{1}{1-\varphi}}}{\partial X_i} = \frac{\partial P_j}{\partial X_i}\frac{\mathbb{P}_j}{\mathbb{P}_j^2} - \frac{\partial \mathbb{P}_j}{\partial X_i}\frac{P_j}{\mathbb{P}_j^2} \\
&= \frac{\partial P_j}{\partial X_i}\frac{1}{\mathbb{P}_j} - \frac{P_j}{\mathbb{P}_j}\sum_{i'}\frac{1}{P_{i'}}\xi_{i'j}\frac{\partial P_{i'}}{\partial X_i}
\end{aligned}$$

I compute the matrices $\left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{P}}\right]$ and $\left[\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{X}}\right]$ numerically with the forward finite difference method.

C Counterfactual Appendix

C.1 Jumpstarting America

Gruber and Johnson (2019) identify potential technology hubs as cities with "sufficiently high population, educational attainment, and quality of life." The list of hubs is available at https://www.jump-startingamerica.com/102-places-for-jumpstarting-america. I select the top 27 cities from that index, which include some dyads and triads and so corresponds to 20 cities or groups of cities. A balanced-budget research subsidy of the same magnitude as CHIPS Act funding corresponds to a 5.4% subsidy in the 27 cities and a -0.8% tax elsewhere. Gruber and Johnson (2019) advocate spending on the order of $100 billion per year but the size of the subsidy does not qualitatively affect the conclusions and so I mimic CHIPS to facilitate the comparison.

The results of this "Jumpstart" program are in Table C1. Aggregate consumption falls very slightly, by about 0.1%. Again the policy is ineffective because it is poorly targeted. Through the lens of the model, the jumpstart cities do not have inefficiently little innovation.

Table C1: Jumpstarting America vs optimal policy

City	Optimal subsidy	Log profit (relative)	Wage change under CHIPS (%)
Atlanta-Sandy Springs-Roswell, GA	-0.055	0.000	-0.057
Grand Rapids-Wyoming, MI	-0.057	0.031	-0.044
Albany-Schenectady-Troy, NY	-0.065	0.023	-0.054
Canton-Massillon, OH	-0.086	0.013	-0.054
Syracuse, NY	-0.086	0.019	-0.054
Appleton, WI	-0.097	0.022	-0.054
Oshkosh-Neenah, WI	-0.098	0.066	-0.048
Pittsburgh, PA	-0.099	0.074	-0.055
Des Moines-West Des Moines, IA	-0.103	0.020	-0.054
St. Louis, MO-IL	-0.107	0.099	-0.040
Bloomington, IL	-0.107	0.064	-0.048
Cedar Rapids, IA	-0.108	0.064	-0.052
Akron, OH	-0.116	0.067	-0.048
Champaign-Urbana, IL	-0.121	0.068	-0.048
Binghamton, NY	-0.124	0.070	-0.046
Dayton, OH	-0.127	0.059	-0.046
Iowa City, IA	-0.131	0.061	-0.053
Utica-Rome, NY	-0.142	0.086	-0.039
Dallas-Fort Worth-Arlington, TX	-0.146	0.084	-0.044
Buffalo-Cheektowaga-Niagara Falls, NY	-0.149	0.096	-0.040
Columbus, OH	-0.149	0.097	-0.053
Lafayette-West Lafayette, IN	-0.159	0.108	-0.040
Rochester, NY	-0.164	0.102	-0.041
Cincinnati, OH-KY-IN	-0.165	0.104	-0.038
Indianapolis-Carmel-Anderson, IN	-0.168	0.103	-0.040
Green Bay, WI	-0.170	0.116	-0.043
Cleveland-Elyria, OH	-0.198	0.102	-0.041

"Jumpstarting America" refers to the top candidates from Gruber and Johnson (2019). Second column refers to the optimal subsidy and third column to the log private value V of research, relative to the average city (both averaged across skill groups). Fourth column is wage change from baseline to the policy.

D Occupation-specific patenting

In the empirical analysis of Section 2 I showed that patenting and wages increased with the stock of neighbors' patents at the city level. One potential concern is that I may be simply picking up regional exposure to a common sector-level shock. For example, San Francisco's patenting and wages may appear to increase with San Jose's patenting not because the two cities are geographically close but because they both specialize in information technology. On the flip side, patenting and wages in Detroit may not increase with patenting in San Jose not because the two cities are geographically distant, but because they produce different goods and use different technologies.

To address this concern, I conduct a more granular analysis at the level of the city and occupation. I revise regression (2) to the form

$$\Delta \log Y_{j,o,t} = \phi^Y \sum_{i \neq j} \frac{\text{distance}_{ij}^{-1}}{\sum_{i' \neq j} \text{distance}_{i'j}^{-1}} \Delta \log \text{patents}_{i,o,t-10} + \Xi' D_{j,o,t} + v_{j,o,t}. \tag{65}$$

where j indexes cities as before and o now indexes two-digit SOC occupations. The right-hand side variable is lagged average patent growth in the city's neighbors, in the same occupation. In Section D.1 I describe how I measure patents at the occupation level. The controls are the lag of patent growth in the own city and occupation as well as city $\times$ year and occupation $\times$ year fixed effects. Exploiting differences only across occupations within a city at a point in time is a demanding cut of the data.

I instrument for log patent growth in the neighbors using lagged log ancestry-weighted changes in foreign countries' occupational patent shares. Namely, the instrument for patent growth in city i and occupation o at time t is

$$\Delta \log \widehat{\text{patents}}_{i,o,t} = \Delta \log \left(\sum_f \text{ancestry share}_{i,f,1900} \times \frac{\text{patents}_{f-US,o,t-10}}{\sum_{o'} \text{patents}_{f-US,o',t-10}} \right) \tag{66}$$

The results are in Table D1. The first stage regression is in Table D2.

I emphasize that the ancestry share is the same not only over time but across occupations with a city. The estimator is therefore consistent even if historical ancestry is correlated with a city-level shock. The exclusion restriction is that the city's ancestry share in 1900 is not correlated with *occupation-specific* foreign patent growth.

Table D1: Effects of neighbors' patents (occupation-specific)

	Δ Log patents$_t$		Δ Log wage$_t$	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
Δ Log neighbors' patents$_{t-10}$	1.20***	4.34***	0.006	0.521***
	(0.104)	(0.632)	(0.020)	(0.145)
Δ Log patents$_{t-10}$	-0.388***	-0.402***	0.0004	-0.002
	(0.007)	(0.008)	(0.001)	(0.001)
City × Year FE	✓	✓	✓	✓
Occupation × Year FE	✓	✓	✓	✓
N	42,182	42,182	41,712	41,712
R^2	0.16	0.13	0.00	-0.01
KP F stat		393.88		363.23

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Occupations are SOC two digit. Bartik instrument interacts country f ancestry share in 1900 with lagged country f patenting by occupation level. Patents assigned to two-digit occupations using natural language processing (NLP) assignment algorithm (see text).

Table D2: Effects of neighbors' patents (occupation-specific), first stage

	Δ Log neighbors' patents$_{t-10}$
	(1) OLS
Δ Log neighbors' patents$_{t-20}$	0.263***
	(0.013)
Δ Log patents$_{t-10}$	0.004***
	(0.0005)
City × Year FE	✓
Occupation × Year FE	✓
N	42,182
R^2	0.03

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$. Standard errors clustered at the city level. Occupations are SOC two digit. Bartik instrument interacts country f ancestry share in 1900 with lagged country f patenting by occupation level. Patents assigned to two-digit occupations using natural language processing (NLP) assignment algorithm (see text).

D.1 Measuring occupation-level patenting

My approach is based on Kogan et al. (2021), and the interested reader is referred to that paper for more detail. I allocate each patent in the data to one or more occupations using modern tools from the Natural Language Processing (NLP) literature. I compare the semantic content of patent text against text from O*NET, an occupational dictionary. Patents are then assigned to the most idiomatically similar occupations.

I use two sets of text, or corpora. The patent corpus is the abstract text of all patents granted by the US Patent and Trademark Office between 1976 and 2019. The occupation corpus is the set of tasks associated to each 6-digit occupation in the Occupational Information Network (O*NET). O*NET is a resource created by the US Department of Labor which describes occupational characteristics and requirements.

First, I encode each patent abstract b and each occupation description c as a vector v in high-dimensional Euclidean space. I use the TensorFlow Universal Sentence Encoder, a general-purpose pre-trained encoder which accounts for context and word order (Cer et al. 2018). This algorithm returns a vector of dimension 512. One can think of these 512 dimensions as representing the most important components of language. Arithmetic operations on vectorized text is meaningful. For example, consider a classic example from the NLP literature:

$$v_{\text{king}} - v_{\text{man}} + v_{\text{woman}} \approx v_{\text{queen}}$$

I can then define the similarity index of any two strings as the inner product of their vector representations. For example, $v_{\text{king}} \cdot v_{\text{serf}} < v_{\text{peasant}} \cdot v_{\text{serf}}$, consistent with human intuition about how related these words are. The encoder works on sentences and paragraphs too. The strings "How old are you?" and "What is your age?" score as virtually equivalent, despite having no exact substrings in common.

I then compute the similarity index of every patent and every six-digit SOC occupations—roughly seven billion comparisons. Each index is a number between 0 and 1, where 0 corresponds to essentially uncorrelated texts and 1 means the texts are identical. For each patent, I retain the five six-digit SOC occupations with the highest similarity indices, $\text{similarity}^i_{b,c}$ for $i = 1, \ldots, 5$. These top five are the match scores.[29] I then compute the empirical cumulative distribution function, $\widehat{F}$, of these match scores, so $\widehat{F}(s)$ is the share of retained scores lower than s. The normalized match score is $\widehat{F}(\text{similarity}^i_{b,c})$. The purpose of the normalization is to infer how good a match is from the observed frequency of match

[29] Four patents—out of over seven million—registered negative similarity scores for their fifth best matching occupation. I recoded these to zero.

scores. Finally, I compute the implied number of patents in every city, occupation, and time period by adding up $\widehat{F}(\text{similarity}^{i}_{b,c})$ over the relevant abstracts b in that city and time, and over the six-digit occupations c in each two-digit occupation.

I provide a couple of examples of how the match algorithm works:

Example. Patent 5,324,077: "Medical Data Draft for Tracking and Evaluating Medical Treatment." The patent describes a system to collect and transmit medical information to insurers. The best-matching occupation for this patent is 31-9092, medical assistants. One of the core tasks performed by medical assistants is "Record patients' medical history, vital statistics, or information such as test results in medical records." The other four occupations receiving positive weight from this patent are 29-1171 (nurse practictioners), 29-1071 (physician assistants), 31-9094 (medical transcriptionists), and 29-2032 (diagnostic medical sonographers). This patent receives an overall weight of 60% for SOC-29 and 40% for SOC-31.

Example. Patent 5,960,411: "Method and system for placing a purchase order via a communications network." This is the patent for Amazon's one-click buy. The patent has 60% weight to SOC-43 (Office/Admin), 20% to SOC-41 (Sales), and 20% to SOC-13 (Business/Financial). Key O*NET tasks associated with occupations within SOC-43 are "Prepare purchase orders" and "Compare prices, specifications, and delivery dates." Although this is a software patent, it is not assigned to software occuaptions.

Example. Patent 8,767,980: "Omnidirectional button-style microphone." This patent has 87% weight on SOC-27 (Arts/Design/Entertainment/Media), with a small amount of weight on SOC-51 (Production) and SOC-23 (Legal). The relevant O*NET task within SOC-27 is "Control audio equipment to regulate volume and sound quality." Although this is a hardware patent, it is not assigned to occupations producing hardware.

Figure D1 shows, for each occupation, the share of total patents with positive weight in that occupation. The figure is pooled over the full sample. The numbers add to more than 100% because each patent can match to multiple occupations. Note that the overall scale is not relevant in the regression because I always control for occupation $\times$ year fixed effects, so that I only exploit variation within an occupation and year.

Figure D1: Share of patents receiving positive allocation weight, by occupation