R&D Subsidies and the Geography of Innovation*

Felipe Camêlo and Alexandre B. Sollaci†

International Monetary Fund

May 13, 2025

Abstract

We study the impacts of the spatial distribution of R&D subsidies in the US. One key feature of these subsidies is their dynamic effects (through innovation, as they subsidize R&D) and spatial effects (as the subsidy rates vary from state to state). To account for both dimensions, we develop a tractable model of endogenous growth with spatial heterogeneity and agglomeration spillovers in innovation. We provide three key results. First, the existing distribution of R&D subsidies is superior to a spatially homogeneous one with the same overall cost. Second, the optimal distribution of R&D subsidies increases the present value of consumption by 3.2% to 6%, without increasing the overall fiscal burden of the subsidy. Third, the geographical scope of the policy (e.g., subsidizing cities vs. states) can drastically change which locations get subsidized.

Keywords: Innovation, firm dynamics, agglomeration spillovers, R&D subsidies, spatial policies

*A previous version of this paper was circulated as "Agglomeration, Innovation, and Spatial Reallocation: the Aggregate Effects of R&D Tax Credits." We would like to thank Rodrigo Adão, Ufuk Akcigit, Jonathan Dingel, Antonio Gabriel, John Grigsby, Krisztina Orban, Stefano Pegoraro, Esteban Rossi-Hansberg, Chad Syverson, James Traina, as well as participants at several seminars and conferences for helpful comments and discussions. The views expressed in this paper do not necessarily represent those of the IMF, its Executive Board, or IMF management.

†e-mail: Camêlo: `fcamelo@imf.org`; Sollaci: `abalduinosollaci@imf.org`

1 Introduction

Research and Development (R&D) subsidies are some of the most important policies to foster innovation in the US. However, these incentives often vary substantially between locations. Consider R&D tax credits: Introduced in 1981 at the federal level, these credits apply to company-funded R&D expenditures above a predetermined baseline level.[1] Since their introduction, most states have also adopted subsidies targeting research activities over the years (see appendix figures A.1 and A.2), resulting in a wide spatial dispersion of R&D tax credits. Figure 1 plots the dispersion of effective tax credit rates by state in 2005.

What are the aggregate implications of the spatial dispersion in R&D incentives? The answer to this question must account for both the spatial *and* dynamic implications of R&D policy. Spatially heterogeneous subsidies can affect where workers and firms choose to locate (Moretti and Wilson, 2017; Akcigit et al., 2016). This is relevant because innovative activity benefits from agglomeration spillovers: inventors are more productive when they are located in densely populated cities (Moretti, 2021).[2] Moreover, as R&D constitutes an investment decision, it affects not only the level of economic activity but also economic growth.

In contrast, most existing studies that evaluate the impact of spatial policies do so in a static setting (Kline and Moretti, 2014; Ossa, 2015; Gaubert, 2018; Fajgelbaum and Gaubert, 2020). Similarly, studies that evaluate the long-run effects of R&D or innovation policies seldom account for local externalities or the spatial heterogeneity of innovation subsidies (e.g., Akcigit et al., 2018).

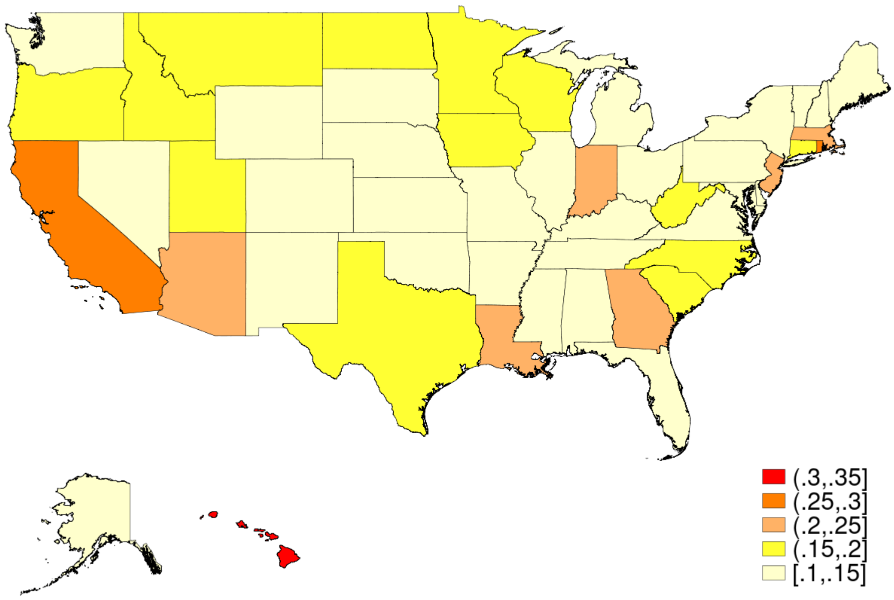
We build an endogenous growth model with local agglomeration externalities and spatial heterogeneity to assess the current distribution of R&D subsidies and compute a potential optimum. The framework captures the spatial and dynamic aspects of R&D policy by allowing the productivity of individual inventors to increase with the population density of the city where they live. Furthermore, the population distribution itself is endogenous, which means that changes in local policies can attract new inventors to a city. We provide three main findings. First, the current distribution of R&D subsidies does better than a spatially homogeneous subsidy that would cost the same to the general government, increasing welfare (present value of consumption) by about 1%. Second, there is much to be gained under an optimal policy. If subsidies are allowed to vary by city, aggregate welfare increases by at least 6% under the optimal distribution. When subsidies are fixed at the state level, the gains are about half as large, at 3.2%. In both cases the overall fiscal burden is kept constant, so these gains are driven solely by the reallocation of inventors to a few dynamic cities.

Third, and perhaps most surprising, the geographical scope of the policy (e.g., city vs. statewide subsidy) plays a critical role in determining which locations should receive subsidies. Since agglomeration externalities operate at the city level in the model, concentrating inventors in a few highly productive cities will boost innovation. At the state level, this idea may not hold because the composition of cities also plays a role. Specifically, it is sub-optimal to subsidize

[1] The Joint Committee on Taxation (2010) estimated that R&D tax credits would amount to over US$ 9 billion in foregone revenues to the federal government in 2019, about 20% more than the National Science Foundation's budget for that year (see `https://www.nsf.gov/about/budget/fy2019/pdf/01_fy2019.pdf`).

[2] The benefits from agglomeration, however, must be weighed against congestion costs and the fact that local benefits might be offset by losses elsewhere (see Kline and Moretti, 2014).

Figure 1: R&D Tax Credit Rates per State, 2005.



Note: The figure shows the *effective* R&D tax credit rates as computed by Wilson (2009). Statutory and effective credit rates differ based on how the base amount is defined and whether or not the credit is "recaptured" (i.e., considered taxable income).

R&D in states that have one highly productive city *and* several other medium-sized cities. This is because the subsidy to these medium-sized cities will also attract inventors, undermining the concentration into the most productive places. In contrast, a state where innovation is highly concentrated in a single city is a good candidate for a state-level subsidy—even if that city is not the most productive in the country—because most inventors will end up in the same place.

In its essence, the framework we develop adds a spatial dimension to a model of endogenous growth through innovation (Aghion and Howitt, 1992). The model includes a system of heterogeneous cities, where inventors benefit from agglomeration spillovers (i.e., they become more productive when they are in densely populated by other inventors), but may also face congestion costs through more expensive non-tradable goods. Firms hire inventors to innovate over goods varieties, increasing their quality and pricing competitors out. As a result, firms want to be located in the most productive places (including agglomeration spillovers), which allows them to innovate more and increase profit. However, because workers can move across cities after paying a moving cost, this force for agglomeration is countervailed by higher wages required to live in more expensive cities.

This is where government policies can play a role in determining the spatial distribution of innovation. R&D subsidies reduce firms' innovation costs, increasing their innovation efforts. If these subsidies are local and focused on densely populated regions, they also increase the willingness of firms to pay the higher wages required to live in those places. Alternatively, if subsidies are focused on low-density areas, they incentivize firms to move away from large population centers. As a result, governments can use local policies to influence the location choices of inventors and firms, taking advantage of agglomeration spillovers to increase aggregate innovation or reducing the costs of R&D through lower congestion.

Despite these moving parts, the model's equilibrium is highly tractable. This makes the policy exercises we run computationally simple, as many key variables have analytical solutions in equilibrium. More importantly, it provides us with relationships between variables that can be directly taken to the data, making identifying and estimating parameters more transparent. For example, two key parameters in the model are the elasticity of agglomeration and the elasticity of congestion with respect to city size. Within the model's structure, these parameters can be estimated through linear regressions that we take to our data. To account for potential endogeneity in each city's population, we construct a shift-share instrument that leverages industry-specific growth in the employment shares of inventors as exogenous shifters for the population in each city (Adão et al., 2019; Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020).

The model also does well in replicating untargeted distributions of patents and firms and in replicating out-of-sample changes in the spatial distribution of inventors and patents over several decades. Specifically, the model can reproduce the cross-city distribution of inventors and patents in the 1970s, 1980s, and 1990s with reasonably high accuracy (correlations greater than 0.8) just by changing the R&D subsidy to its appropriate level at the time. We note that changes in state-specific R&D tax credits/subsidies are largely uncorrelated with changes in states' corporate and labor income taxes, which suggests that R&D subsidies are indeed responsible for the long-run patterns we measure.

1.1 Relationship to the Literature

This paper is related to several strands of the economic literature. The main findings of the paper contribute to the literature on spatial misallocation and optimal spatial policies. In line with our results, several other studies have found large potential gains from reallocating resources across space in the US.[3] Hsieh and Moretti (2019) argue that housing supply restrictions adopted in some of the most productive cities in the US significantly lowered the country's rate of growth between 1964 and 2009. Fajgelbaum et al. (2019) find that tax dispersion across states leads to aggregate losses because it distorts the spatial allocation of resources. Gaubert (2018) and Fajgelbaum and Gaubert (2020) develop general quantitative frameworks that allow them to compute optimal local subsidies designed to attract workers and firms to each city. Similarly, Ossa (2015) explores the welfare effects of subsidy competition among states, while Kline and Moretti (2014) study the long-run effects of the Tennessee Valley Authority development program.

Although these are important contributions, the existing literature has analyzed the welfare effects of spatial policies primarily through static frameworks. As discussed above, R&D tax credits have both spatial and dynamic effects that interact with each other. A purely spatial model would incorrectly evaluate the welfare effects of spatially reallocating R&D subsidies

[3] In light of the recent rise in economic inequality across the US, some economists have diverted their focus to policies designed to mitigate economic distress (Austin et al., 2018) or income inequality Farrokhi (2021). Glaeser and Hausman (2019) discuss changes in innovation policies that focus on particularly impoverished areas as a means of reducing joblessness. At the same time, they recognize that spatially reallocating innovative activity might reduce the overall production of innovation in the country. So, they suggest alternatives that minimize the spatial reallocation of R&D funding.

because it would ignore the impact of a rise/fall in the rate of creative destruction on the incentives of firms (more creative destruction reduces the incentives to innovate, as the chance that a firm is replaced by a competitor in the future increases). Similarly, a purely dynamic model would predict no relationship between the location of subsidies and aggregate growth. Thus, to understand how the aggregate economy reacts to changes in the distribution of R&D subsidies, one must adopt a setting that captures this policy's dynamic and spatial effects.

The theoretical framework developed here is based on endogenous growth models in which innovation through creative destruction is the main driving force of economic growth (Aghion and Howitt, 1992; Klette and Kortum, 2004; Akcigit and Kerr, 2018). It contributes to this literature by nesting such an endogenous growth model into a spatial setting. The resulting model retains many of the features in the growth literature, but also allows for spatial heterogeneity, agglomeration spillovers, and endogenous population distribution. Adding a spatial dimension to this class of models helps us to understand the linkages between innovation at the firm level, which is greatly impacted by the firm's location through agglomeration spillovers, and broader economic growth. It is also fundamental to the investigation of the impact of spatial policies on aggregate welfare and growth.

This framework also contributes to the literature on spatial and dynamic models. One closely related paper is Duranton (2007), which embeds a quality ladder model into an urban structure to study the city size distribution and the movement of industries across cities. The main advantage of the current setting is that agglomeration and congestion externalities are microfounded and endogenously generated in the model's equilibrium; in contrast, Duranton (2007) relies on a reduced form that captures the net effect of these externalities on the size of cities.

Other spatial and dynamic models (Desmet et al., 2018; Caliendo et al., 2019) generally do not focus on the details of innovation and their solutions and estimation can be numerically demanding. In contrast, our model remains highly tractable and preserves a natural link between firm-level decisions in the model and in the data. This tractability is partly possible because some frictions, such as trade costs, are ignored in our model as they typically do not account for a significant cost of R&D-intensive products/industries. Other relevant spatial heterogeneities (amenities, moving costs, city-specific productivity) are also crucial in our framework.

The remainder of this paper is organized as follows. Section 2 develops a formal endogenous growth model replicating the linkages between city-specific R&D subsidies and aggregate growth. Section 3 solves for the model's equilibrium in a Balanced Growth Path. Section 4 shows how to map the model's equilibrium to the US data and estimate its parameters. In section 5, we use this framework to measure the effects of adopting alternative R&D policies. Section 6 concludes.

2 A Dynamic Spatial Model of Innovation and Growth

In this section, we introduce our theoretical framework. The main contribution of this model is to provide a tractable quantitative framework that allows the productivity of R&D investments to be a function of spatial heterogeneity and local agglomeration externalities. This implies that the spatial distribution of economic activity matters for aggregate growth. In addition, since workers and firms choose where to locate, this distribution is endogenous and responds dynamically to policy changes.

2.1 Cities

There are $C+1$ cities, indexed by $c \in \{0, 1, \dots, C\}$. For many results in the paper, we consider the limit as $C \to \infty$. Cities differ in four key aspects. First, each city has an amenity level a_c that captures factors enhancing quality of life, such as climate, geography, or cultural attributes. Second, cities have a landmass $\bar{m}_c$ which absentee landlords own. For convenience, we assume land is fully taxed by the government and normalize $\bar{m}_0 = 1$ and $\bar{m}_c = 1/C$ for all $c \geq 1$. This emphasizes that when C is large, each city represents only a small share of the economy, and the Law of Large Numbers can be applied to equilibrium results.[4] The distinct role of city 0 will be discussed in below.

Third, cities have a distinct geographical location. The impact of a city's location on economic outcomes can be felt through moving costs: any individual who wishes to move from city c to city c' must pay an instantaneous cost that reduces their utility by a factor of $d(c, c')$. Following Desmet et al. (2018), we assume that this cost can be decomposed into an origin- and a destination-specific component such that

$$d(c, c') = d_1(c)d_2(c') \quad \text{and} \quad d(c, c) = 1,$$

where the latter condition guarantees that no cost is incurred if the agent decides not to move. These two assumptions imply that

$$d_1(c) = \frac{1}{d_2(c)}$$

for any city c, which means that an agent leaving city c effectively receives a "benefit" equal to the cost of entering that same city. One important implication of this result is that the history of cities where an agent has lived does not affect their overall lifetime moving cost: this is fully determined by the initial and final destinations only. To see why, note that the cost of moving from city a to city b and later from city b to city c is the same as moving from city a to city c directly:

$$d(a, b) \times d(b, c) = d_1(a) \underbrace{d_2(b) \times d_1(b)}_{=1} d_2(c) = d_1(a)d_2(c) = d(a, c).$$

Fourth, and finally, we also introduce a city-specific productivity term $\chi_c(t)$, which captures the idea that some cities serve as innovation hubs, often benefiting from local venture capital networks (Gompers and Lerner, 2001) or a culture favorable to entrepreneurship (Manso, 2011; Saxenian, 1994). To allow for local shocks and realistic spatial dynamics, we assume this productivity term is stochastic and evolves according to

$$\chi_c(t) = \bar{\chi}_c e^{z_c(t)},$$

[4] This normalization is also without loss of generality, as the role of landmass cannot be distinguished from amenities and productivity in equilibrium outcomes.

where $\bar{\chi}_c$ captures permanent differences in R&D productivity across cities, and $z_c(t)$ is an Ornstein-Uhlenbeck (O-U) process:

$$dz_c(t) = \phi\big(\mu - z_c(t)\big)dt + \sigma\, dW_c(t).$$

Here, ϕ and σ denote the mean reversion and volatility parameters, respectively, while $W_c(t)$ is a city-specific Brownian motion. The O-U process is particularly useful in this context as it ensures stationarity, which means that as $t \to \infty$, the distribution of $z_c(t)$ stabilizes with a finite mean and variance (Stokey, 2008). Consequently, while individual cities face uncertainty, the model does not exhibit aggregate uncertainty in the large-C limit. We set $\mu = -\sigma^2/(4\phi)$ so that, under the stationary distribution, $\mathbb{E}[e^{z_c}] = 1$.

People and Goods. Each city hosts both workers and firms. Workers are divided into two groups: inventors, who engage in R&D, and production workers, who manufacture final goods. The total number of inventors and production workers, denoted by I and L respectively, remains constant over time. Workers do not switch occupations in response to R&D subsidies, an assumption that simplifies the model and aligns with empirical findings on the inelasticity of inventor supply (Goolsbee, 1998).

Both types of workers are free to move across cities at any time (after paying the moving costs) but must be employed locally—meaning that they work in the city where they reside. This assumption might also abstract from some real-world frictions but aligns with empirical evidence that labor responds significantly to regional tax differentials (Moretti and Wilson, 2017). Firms can also choose their location upon entry but cannot relocate afterward. In our model, the location of a firm determines where its innovation activities occur. Firms with multiple locations that employ inventors in different cities are treated as separate entities (see Section 4 and Online Appendix C.1). Like land, firms are owned by absentee investors and are fully taxed by the government.

The model includes three types of goods: a final non-tradable good, a final tradable good, and intermediate goods. We assume that tradable goods can be exchanged without transportation costs. Although this is a simplification, it is justified by the fact that many innovation-intensive goods have low transport costs relative to their value. Additionally, production and innovation often occur in different locations, implying that transport costs should not significantly influence firms' decisions on where to innovate. In our setup, production is concentrated in city 0, whereas innovation occurs in all other cities.

2.2 Preferences

The only consumers in this model are inventors and production workers. All consumers share the same utility function but may differ in wages. A worker of type $h \in \{i, \ell\}$ (where i denotes an inventor and ℓ denotes a production worker) has utility given by

$$U^h(t_0) = \int_{t_0}^{\infty} e^{-\rho t} \max_{c(t)} \{u^h_{c(t)}(t) G(t)\} dt,$$

where $c(t)$ is the city where the worker resides at time t, $G(t)$ is the amount of public good consumed in period t, and $u^h_{c(t)}(t)$ represents the maximum utility a worker of type h can attain while living in city c at time t:

$$u^h_c(t) = \max_{n(t),y(t)} [a_c n(t)]^\theta y(t)^{1-\theta} \quad \text{s.t.} \quad y(t) + p_{n,c}(t)n(t) \leq w^h_c(t).$$

In this expression, a_c are amenities, $n(t)$ represents the amount of the non-tradable final good consumed at time t, and $y(t)$ is the amount of the tradable final good consumed in the same period. In what follows, we will refer to the non-tradable final good simply as the "non-tradable good" and the tradable final good as the "final good" (intermediate goods are introduced below).

All workers inelastically supply one unit of labor per period, so their income is equal to their wage, $w^h_c(t)$. To simplify the analysis, we assume that all consumers are "hand-to-mouth" in the sense that they cannot borrow or save. This ensures that a worker's wealth does not depend on past locations where they may have lived.

The Role of Mobility Restrictions. Appendix G explores several extensions of the model, one of which assumes no mobility across cities. In this scenario, we show that the spatial distribution of R&D tax credits does not affect the economy's aggregate innovation rate. Any impact of larger subsidies is entirely absorbed into higher wages, as the supply of inventors in each city remains fixed. This result contradicts empirical evidence that R&D subsidies increase innovation at the firm level and that inventors exhibit relatively high mobility.

An alternative extension assumes that inventors can move freely across cities, while production workers cannot. In this case, the incentive for inventors to agglomerate becomes even stronger, as they continue to benefit from localized knowledge spillovers and network effects while facing lower congestion costs.

2.3 Technology

The production side of the economy consists of three sectors: non-tradable goods, final goods, and intermediate goods. Non-tradable goods are locally produced using land and labor, while final goods are produced using labor and intermediate goods. Intermediate goods, in turn, are produced by firms that innovate and accumulate product lines over time. Firms compete in a pricing game, and the technological leader in each variety effectively acts as a monopolist.

2.3.1 Non-tradable Good

Non-tradable goods are locally produced by competitive firms using land and the labor of production workers. The non-tradable goods producer in each city solves

$$\max_{\ell_{n,c}(t),m_c(t)} p_{n,c}(t)n(t) - w^\ell_c(t)\ell_{n,c}(t) - p_{m,c}(t)m_c(t) \quad \text{s.t.} \quad n(t) = \ell_{n,c}(t)^\beta m_c(t)^{1-\beta},$$

by hiring $\ell_{n,c}$ production workers and renting m_c units of land. The producer takes as given the price of the non-tradable good, $p_{n,c}$, the price of land, $p_{m,c}$, and the wages of production workers, w_c^ℓ.

Land is a fixed factor in each city, leading to decreasing returns to scale in the production of non-tradable goods. This generates congestion costs: as a city's population grows, so does the demand for the non-tradable good. Due to decreasing returns to scale, higher demand pushes up prices, making the city less attractive to live in.[5]

2.3.2 Final Goods

The final good is produced competitively using labor and intermediate goods. Like all firms, final good producers are free to choose their location at entry. If production occurs in city c^*, the final good producer solves

$$\max_{\ell_{y,c^*}(t),\{k_j(t)\}_{j\in\mathcal{J}}} Y(t) - \int_{\mathcal{J}} p_j(t)k_j(t)dj - w_{c^*}^\ell(t)\ell_{y,c^*}(t) \quad \text{s.t.} \quad Y(t) = \frac{\ell_{y,c^*}(t)^\varepsilon}{1-\varepsilon}\int_{\mathcal{J}} k_j(t)^{1-\varepsilon}q_j(t)^\varepsilon dj,$$

where ℓ_{y,c^*} is the number of production workers hired, and k_j and q_j denote the quantity and quality of intermediate good j used in production, respectively. Each intermediate good is sold at a price p_j to final good producers, and the price of the final good is normalized to 1.

Since all intermediate goods and the final good are freely tradable across cities, their production location does not affect their prices.

2.3.3 Intermediate Goods

Intermediate goods exist in a continuum of varieties, $j \in \mathcal{J} \equiv [0,1]$. Intermediate goods producers—hereafter simply "firms"—can produce multiple varieties. Each firm's product set and the quality at which it can produce them consists of all innovations it has previously developed. For now, we take this set as given.

Following Akcigit and Kerr (2018), firms compete in a two-stage pricing game:[6]

Assumption 1. *All firms capable of producing intermediate good j engage in a two-stage pricing game. In the first stage, firms decide whether to pay a small fee to announce their price. In the second stage, all firms that paid the fee set their selling price for good j.*

This assumption directly implies that only the firm producing good j at the highest quality (the technological leader) pays the fee and enters the second stage. As shown in equilibrium, demand for intermediate goods increases with quality, so technological laggards cannot recoup the fee if they enter the second stage. Consequently, the technological leader acts as a monopolist in producing good j.

A crucial requirement for this result is that firms face identical production costs regardless of location. This rules out production functions where labor is used to produce intermediate goods, as wages vary across cities. If location affected production costs, the technological leader might

[5] The production function implies that the elasticity of the supply of non-tradable goods with respect to population is the same across cities. This assumption abstains from different legal environments across cities, for example, in the stringency of zoning laws or the ease of building housing. Empirically, Hsieh and Moretti (2019) document variation in elasticity across cities, which may impact labor allocation.

[6] This pricing game leads to an outcome similar to limit-pricing, except that firms' markups depend on their production elasticity ε rather than the quality step size λ.

be unable to drive lower-quality competitors out of the market. For instance, firms producing at lower quality but facing lower costs could remain competitive by undercutting the leader's price.[7]

Following Akcigit et al. (2018), we assume that the final good is used as an input in the production of intermediate goods. Since the final good is freely tradable, all firms face the same marginal cost $\nu > 0$. The technological leader in each variety chooses production by solving:

$$\max_{k_j(t)} p_j(k_j(t); q_j(t))k_j(t) - \nu k_j(t),$$

where $p_j(k_j; q_j)$ represents the inverse demand function for intermediate good j, and the price of the final good is normalized to 1 in all periods.

An important implication of this framework is that intermediate goods production is not city-specific. This makes the model equivalent to one in which production and innovation are geographically dissociated, with production occurring wherever it is cheapest. The only incentive for firms to locate in a particular city stems from location-specific productivity differences in R&D investments, which ultimately determine the set of products a firm can produce.

2.4 Research and Development

Innovation in this model plays two key roles. First, it drives economic growth by improving the quality of intermediate goods. Second, it expands firms' product portfolios by establishing them as the technological leader in a given variety of intermediate goods. When a firm f innovates in product line j, it immediately produces that good at quality $(1+\lambda)q_j(t)$, where $\lambda > 0$ represents the quality improvement from innovation (or the *step size* in the quality ladder), and $q_j(t)$ is the highest quality at which good j is currently produced. Since the innovating firm becomes the new leader for good j, it displaces the previous producer, generating creative destruction.

Firms invest in R&D to generate innovations. Following Klette and Kortum (2004), the number of innovations realized in each period follows a Poisson process, which in continuous time implies that firms produce at most one innovation per period.[8] Firms cannot target specific product lines with their R&D investments, meaning that any given innovation occurs randomly across all product lines in the economy.

This assumption has several implications. Since firms cannot deliberately target competitors' products, R&D investment decisions are independent of competitors' actions. Additionally, because firms start with no product lines and can only acquire them through innovation, the number of products in any firm's portfolio remains countable. With a continuum of product lines in the economy, the probability that a firm innovates over one of its own products is zero. However, firms must still consider the probability of losing products they currently produce due to competitors' innovations. Another consequence is that firms do not belong to specific sectors; they can produce any good they innovate upon. While common in endogenous growth

[7] A similar issue arises if there are transportation costs for intermediate goods, as in Eaton and Kortum (2002). In such cases, less productive firms may still sell goods in nearby markets, meaning the most technologically advanced firms do not necessarily dominate sales.

[8] Fix x as the arrival rate of an innovation for an arbitrary firm. The number of arrivals N in a period of length Δ follows $\mathbb{P}(N = k) = e^{-\Delta x}(\Delta x)^k/k!$. Expanding $e^{\Delta x}$ in a Taylor series, we obtain $\mathbb{P}(N = 0) = 1 - \Delta x + o(\Delta)$, $\mathbb{P}(N = 1) = \Delta x + o(\Delta)$, and $\mathbb{P}(N \geq 2) = o(\Delta)$.

models, this assumption precludes sectoral specialization across cities, a feature often observed in practice. This likely reduces the benefits from geographic concentration, as inventors interact across sectors rather than primarily within their own.

Innovation depends on both firm-level R&D investment and the city in which the firm operates. Innovation requires hiring inventors, who benefit from agglomeration externalities in their city. Letting $\tilde{I}_c$ denote the density of inventors per unit of land in city c, the productivity of each individual inventor is proportional to $\tilde{I}_c^{\eta}$, where $\eta \geq 0$ determines the strength of agglomeration spillovers. A firm f located in city c that hires $i_{f,c}$ inventors generates innovations at the arrival rate:

$$x_{f,c}(t) = \chi_c(t) \left(\tilde{I}_c(t)^{\eta} i_{f,c}(t) \right)^{\psi}, \tag{1}$$

where $\chi_c(t)$ is the city's stochastic productivity term. When $\eta = 0$, there are no agglomeration externalities.

Investment in R&D also involves a fixed cost, paid in terms of inventor labor. Each firm that chooses to invest in R&D must hire $\kappa > 0$ inventors to cover managerial and maintenance costs associated with research. This fixed cost serves as a benchmark for the value of firm entry, ensuring that free entry drives the expected value of entrants to zero.[9]

Several assumptions underlie this framework. The model takes inventor density as the key measure of agglomeration externalities, which is empirically validated in Appendix C.3.3. It also assumes that innovation does not scale with firm size, meaning that all firms in the same city innovate at the same rate. An extension in Appendix G explores the case where innovation rates vary with firm size and shows that the model's aggregate predictions remain unchanged.

Agglomeration spillovers are assumed to be homogeneous within cities and nonexistent beyond city borders. In practice, knowledge spillovers likely decay with distance (Rosenthal and Strange, 2001), but the model captures their aggregate effect within a geographic area corresponding to the size of an average city. Another simplifying assumption is that inventors and firms are homogeneous, implying no sorting across cities. If sorting were allowed, differences in productivity across cities would be reflected in the parameter $\bar{\chi}_c$, although sorting itself might respond to policy changes. The implications of sorting for optimal R&D subsidies are unclear. If more productive inventors benefit more from agglomeration, the economy might favor stronger spatial concentration of inventors. However, lower-productivity inventors migrating to larger cities could reduce average productivity in those locations.

Finally, the model assumes that firms cannot improve the quality of products they already own, meaning that all innovation occurs through creative destruction. As a result, firms do not benefit from their past innovations when improving the quality of goods they produce. Since firms in more productive cities innovate more frequently, the absence of internal innovation (see Akcigit and Kerr, 2018) reduces their incentives to locate in such cities, as the risk of losing products to competitors increases.

City 0. To reflect that some US cities have never produced a single patent, we allow for one city where $\bar{\chi}_0 = 0$. City 0 is representative of all cities that do not have minimal necessary

[9] A fixed cost, rather than an entry cost, is useful in this setting as it makes entrants and incumbents symmetric.

conditions for investment in R&D (and therefore do not innovate) in the country. However, companies located in the city 0, can still participate in the production of goods.

2.5 Government: Aggregate and Local Policies

The primary policy instruments analyzed in this paper are local R&D tax credits, which we refer to interchangeably as subsidies for simplicity. These subsidies reduce the effective cost of R&D by transferring a fraction of each firm's research expenditures back to the firm. The subsidy rate, denoted by s_c, is city-specific, meaning that the level of support varies by location. In addition to R&D subsidies, the government provides a national public good, G, which all workers consume.

To finance these expenditures, the government fully taxes firm- and land-owners. In this setting, the provision of the public good serves as a mechanism to redistribute profits and land rents back to workers. In counterfactual exercises, public good expenditures are adjusted to balance the government's budget constraint, while the values of R&D subsidies are taken directly from the data. The amount of public good available in each period is fixed at $G(t) = \bar{G}$ for all t.

Corporate and labor income taxes. Local governments impose various other taxes and subsidies that are not explicitly modeled. Two particularly relevant taxes in this context are corporate and labor income taxes. In the framework of the model, regional differences in these taxes primarily affect the attractiveness of cities to workers and firms. If tax rates remain constant over time, the effect of corporate income taxes can be captured by differences in $\bar{\chi}_c$, while labor income taxes affect a_c.[10]

As shown later, the distribution of inventors across cities depends not only on the R&D subsidy s_c but also on productivity, amenities, and moving costs. Consequently, changes in regional tax policies, such as income tax adjustments, would influence the spatial distribution of inventors and, by extension, the economy's aggregate growth rate. This paper focuses on R&D subsidies, as their primary objective is to increase investment in innovation. Other taxes often serve redistributive functions that fall outside the scope of this analysis (see Section 5.4).

One potential concern is that corporate and labor income tax rates may be correlated with changes in local R&D subsidies. If this correlation exists, it could lead to an overestimation of the impact of R&D subsidies in the data. We address this issue in Section 4.4, where we show that the evolution of R&D tax credits is largely uncorrelated with other local tax policies.

3 The Balanced Growth Path Equilibrium

In this section, we explicitly pose all the problems consumers and firms face and solve for the model's equilibrium. In what follows, we impose two restrictions on the equilibrium. First, we will solve for a Balanced Growth Path, in which all aggregate variables grow a constant rate. Second, we require that each one of the local shocks $z_c(t)$ follows its stationary distribution. We

[10]Labor income taxes also influence the relative prices of tradable and non-tradable goods. Since non-tradable goods are produced and consumed locally, their prices adjust to local conditions. In contrast, tradable goods are consumed across all regions, meaning their equilibrium prices respond to aggregate economic conditions.

refer to this equilibrium as a Stationary Balance Growth Path, formally defined in appendix B. Despite the many moving part of the model, we show that its equilibrium is extremely tractable and can be almost entirely solved analytically.

3.1 The Firm's Static Problem

Each firm's problem can be divided into a static problem and a dynamic problem. In the static problem, the firm chooses how much to produce of each good, considering the set of products it can produce and their respective qualities. In the dynamic problem, the firm chooses how much to invest in R&D after observing the local productivity shock.

Final goods production. The final good producer is free to choose in which city to locate. Because agglomeration externalities do not have any effect over the production of the final good, production will take place in whichever city has the lowest wages (least congestion): city 0. Since $\chi_0(t) = 0$ for all t, no firms investing in R&D will be located there, meaning there are no inventors living city 0. As a result, the only workers in city 0 are the ones hired by the final good producer and those who produce non-tradable goods. In each period, the final good producer's profit maximization problem is (we drop time from the notation as it causes no confusion)

$$\max_{\ell_{y,0},\{k_j\}_{j\in\mathcal{J}}} \frac{\ell_{y,0}^{\varepsilon}}{1-\varepsilon}\int_{\mathcal{J}} k_j^{1-\varepsilon} q_j^{\varepsilon} dj - \int_{\mathcal{J}} p_j k_j dj - w_0^{\ell}\ell_{y,0}.$$

The first-order conditions define the wage in city 0 and the demand for intermediate goods.

Intermediate goods production. Given the demand function (obtained from the problem above), each firm chooses k_j to

$$\max_{k_j} p_j(k_j;q_j)k_j - \nu k_j.$$

Since there are no transport costs and the marginal cost of production does not vary between cities, the firm's production decision is completely independent of its location. Solving the problem above gives the quantity of good j produced, $k_j = q_j\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1}{\varepsilon}}\ell_{y,0}$, and the profit made from each product line $\pi_j = \ell_{y,0}\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}}\varepsilon q_j$. Plugging the expression for k_j into the final good producer's F.O.C. also gives an expression for the wages of production workers in city 0:

$$w_0^{\ell} = \frac{\varepsilon}{1-\varepsilon}\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}} Q, \tag{2}$$

where $Q = \int_{\mathcal{J}} q_j dj$ is the average quality of all intermediate goods produced in the economy.

3.2 Local Wages and Congestion Costs

Before going into the dynamics of firm decisions, it is helpful to understand how local wages respond to a city's population. We start with the observation that because workers can move across cities at any point in time, their current location must be the one where they receive the highest possible utility level after discounting moving costs. Since this condition must hold for all cities, it follows that (see appendix B for the proof):

Lemma 1 (Mobility Equilibrium)**.** *In equilibrium, the indirect utility of a worker of type* $h \in \{i, \ell\}$ *residing in city c satisfies*

$$\frac{u_c^h(t)}{\delta_c} = u^h(t), \quad \forall c, t$$

where $\delta_c := d_1(c)$ *is the cost/benefit of leaving city c, and* $u^h(t)$ *is the equilibrium utility level common across all cities where workers reside.*

In other words, workers' utility levels (adjusted for moving costs) must be the same across all cities and all periods. This condition plays a central role in shaping spatial sorting, highlighting how mobility frictions influence location decisions and the spatial distribution of economic activity. This also implies that local wages must adjust to compensate workers for any variation in utility caused by different levels of amenities, productivity, or congestion costs across cities.

Congestion costs arise because the production of the non-tradable good involves a fixed factor (land), thus displaying decreasing returns to scale. Combined with the fact that the demand for non-tradable goods increases with the population of workers in each city, DRS in the production of non-tradable goods implies that cities with a larger population will also have more expensive non-tradable goods—generating a congestion cost. Thus, as a city's population increases, so does the cost of living. In equilibrium, this means higher wages (due to worker mobility), which means that firms will favor locating elsewhere.

Lemma 2 describes how wages respond to the city's population (see appendix B for the proof). The "tilde" notation indicates variables per unit of land: for example, if I_c is the population of inventors in city c, then $\tilde{I}_c = I_c/\bar{m}_c$. We also define the effective amenity $\alpha_c = a_c/\delta_c^{\frac{1}{\theta}}$, which reflects the amenity level that workers experience while living in city c (i.e., after discounting the moving costs required to arrive there).[11]

Lemma 2. *In equilibrium, wages can be expressed as the product of a congestion-adjusted (baseline) wage and a term that adjusts for congestion costs in each city. Let* I_c *and* L_c*, respectively, as the population of inventors and production workers in each city (with* $I_0 = 0$*). The wages of inventors in cities* $c \in \{1, \ldots, C\}$ *are given by*

$$w_c^i = w^i \left(\frac{\tilde{I}_c^{1-\beta}}{\alpha_c} \right)^{\frac{\theta}{1-\theta}} \quad \textit{where} \quad w^i = \frac{1}{1-\theta} \left[\frac{u^i}{[(1-\theta\beta)]^\theta} \right]^{\frac{1}{1-\theta}} \left(\frac{I}{L - L_0} \right)^{\frac{\theta\beta}{1-\theta}}, \tag{3}$$

while the wages of production workers are

$$w_c^\ell = w^\ell \left(\frac{\tilde{L}_c^{1-\beta}}{\alpha_c} \right)^{\frac{\theta}{1-\theta}} \quad \textit{where} \quad w^\ell = \frac{1}{1-\theta} \left[\frac{u^\ell}{(\theta\beta)^\theta} \right]^{\frac{1}{1-\theta}}, \tag{4}$$

where u^i *and* u^ℓ *are the utility levels of inventors and production workers, respectively. Land*

[11] Note that there is no distinction between workers that choose to move to city c and those that are "born" there. The utility levels in each city are determined in equilibrium by population flows to cities with higher utility levels. Hence, a city with high moving costs δ_c will always have relatively lower utility (all else constant) because fewer workers are attracted to move there.

rents in each of these cities are given by

$$p_{m,c}\bar{m}_c = \frac{(1-\beta)\theta}{1-\theta\beta} w_c^i I_c.$$

In city 0*, the number of production workers hired to produce non-tradable goods is* $\ell_{n,0} = \theta\beta L_0$*, and the number of workers hired to produce the final good is* $\ell_{y,0} = (1-\theta\beta)L_0$*. This gives us a second expression for the wage of production workers, given by*

$$w_0^\ell = w^\ell \left(\frac{\tilde{L}_0^{1-\beta}}{\alpha_0}\right)^{\frac{\theta}{1-\theta}} (\theta\beta)^{\frac{(1-\beta)\theta}{1-\theta}} \tag{5}$$

and the land rent is $p_{m,0}\bar{m}_0 = (1-\beta)\theta w_0^\ell L_0$*.*

Lemma 2 also implies that, in cities where there is innovation, the number of inventors is proportional to the number of production workers: $I_c/I = L_c/(L - L_0)$. As a result, the population of production workers is sufficient to characterize the population of inventors in each city, and vice-versa.

3.3 The Firm's Dynamic Problem

Firms can either be entrants or incumbents. In each period, firms choose how much to invest in innovation, given the set of product varieties they own and the optimal production decisions described above. Incumbent firms take their location as given and are not allowed to move. Entrant firms are free to choose which city to locate in. The timing of decisions in each period is as follows:

(i) The shock $z_c(t)$ is realized and observed in all cities.

(ii) Potential entrants decide whether or not to enter, and in which city to locate.

(iii) Entrants and incumbents decide how many inventors to hire.

(iv) Innovation is realized (based on the arrival rates $x_{f,c}$) and production takes place.

Incumbents The dynamic problem faced by an incumbent firm located in city c can be described by the Hamilton-Jacobi-Bellman (HJB) equation in lemma 3 below. Define $\mathbf{q}_f$ to be the multiset[12] of the qualities of products that the firm is currently producing and D to be the rate of creative destruction of the economy (the aggregate rate of innovation). Because the measure of intermediate good varieties in the economy is one, D also coincides with the probability than any single product line will be "stolen" at any point in time.

Let r be the (exogenous) interest rate and $A = (Q, w^i, D, L_0)$ denote the "aggregate state" of the economy, where Q is the average quality of all intermediate goods produced, w^i is the "baseline" wage of inventors, D is the rate of creative destruction, and L_0 is the population

[12] A multiset is a generalization of a set that allows for multiple instances of each of its elements. The notation $\cup_+$ indicates the multiset union operator, such that $\{a,b\} \cup_+ \{b\} = \{a,b,b\}$. Similarly, $\setminus_-$ indicates the multiset difference operator, such that $\{a,b,b\}\setminus_-\{b\} = \{a,b\}$. $\mathbf{q}_f$ is a multiset because firms can have multiple products that have the same quality level.

of production workers in city 0 (which affects the flow of profits for firms). For convenience of notation, define $\bar{\pi} = (1-\theta\beta)\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}}\varepsilon$ so that the profit per period of each firm is $\pi_j = \bar{\pi}L_0 q_j$ and $Z_c = e^{z_c}$ so that the city-specific productivity in innovation production is $\chi_c = \bar{\chi}_c Z_c$.

Lemma 3. *The problem faced by an incumbent firm located in city $c \in \{1, \ldots, C\}$ is described by*

$$rV_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) - \frac{\partial V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)}{\partial A}\frac{\partial A}{\partial t} =$$

$$\max_{x_{f,c}} \left\{ \begin{array}{l} \sum_{q_j \in \mathbf{q}_f} \bar{\pi} L_0 q_j + x_{f,c}\mathbb{E}_j[V_c(\mathbf{q}_f \cup_+ \{(1+\lambda)q_j\}, \tilde{I}_c, Z_c, A) - V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)] \\ -(1-s_c)w_c^i(i_{f,c}+\kappa) - D\sum_{q_j \in \mathbf{q}_f}[V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) - V_c(\mathbf{q}_f \backslash_- \{q_j\}, \tilde{I}_c, Z_c, A)] + R_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) \\ s.t. \quad x_{f,c} = \bar{\chi}_c Z_c(\tilde{I}_c^{\eta} i_{f,c})^{\psi} \end{array} \right\}$$

The derivation of this equation can be found in appendix B. The first term inside the curly brackets is the profit made through the production and sale of goods. It is followed by the expected gain from one more innovation, which introduces a new product into $\mathbf{q}_f$ (recall that the number of innovations per period follows a Poisson distribution, so that in continuous time the probability that two or more arrivals occur can be ignored). The first term in the second line is the cost of investment in R&D (both variable and fixed), subsidized at rate s_c. The second term in the second line is the expected cost from losing a product line due to creative destruction. The remaining term, $R_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)$, captures the risk that firms in city c face due to the productivity shock.

Corporate income taxes Recall from section 2 that one of the government's revenue sources is taxes on firm owners. Taxing firm owners in this model is the same as taxing a firm's profits, yet no corporate income taxes appear on the HJB equation above. As shown in the proof of lemma 3, this is done because corporate income taxes only shift the share of the firm's value that is accrued to the firm's owner without having any effect on the allocation of resources in the economy. By fully taxing firm owners, we study the limiting case where corporate taxes are arbitrarily close to one but where firms still behave as profit maximizers (as if their value function was V_c in lemma 3).

Entrants Entrants behave the same way incumbents do, with two exceptions: first, entrants do not yet have any product lines of their own; second, they can choose where to locate. As before, define the aggregate state of the economy as $A = (Q, w^i, D, L_0)$. The entrant firm's problem can be solved in two steps:

Step 1: Choose which city to locate in after observing all shocks $\{Z_c\}_{c=1}^{C}$:

$$V^e(A) = \max_c V_c^e(\tilde{I}_c, Z_c, A).$$

Step 2: Choose the level of innovation subject to being in city c.

$$rV_c^e(\tilde{I}_c, Z_c, A) - \frac{\partial V_c^e(\tilde{I}_c, Z_c, A)}{\partial A}\frac{\partial A}{\partial t} =$$

$$\max_{x_{f,c}} \left\{ \begin{array}{l} x_{f,c}\mathbb{E}_j[V_c(q_j, \tilde{I}_c, Z_c, A) - V_c^e(\tilde{I}_c, Z_c, A)] - (1-s_c)w_c^i(i_{f,c}+\kappa) + R_c^e(\tilde{I}_c, Z_c, A) \\ \text{s.t.} \quad x_{f,c} = \bar{\chi}_c Z_c(\tilde{I}_c^{\eta} i_{f,c})^{\psi} \end{array} \right\}$$

The equation for the second stage of the entrant's problem is exactly analogous to the incumbent's problem. Still, it does not include the flow of profits from current production or the risk of losing products to other firms by creative destruction. Proposition 1 describes the value function for entrants and incumbents (see appendix B for the proof).

Proposition 1. *In a Stationary Balanced Growth Path Equilibrium where the production of final goods Y grows at rate $g < r$, the value function of an incumbent firm located in the city $c \geq 1$ and whose portfolio of products is $\mathbf{q}_f$ is*

$$V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = F(D, L_0) \sum_{q_j \in \mathbf{q}_f} q_j + \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)Q\right\},$$

where $F(D, L_0) = \bar{\pi}L_0/(r+D)$ is the "franchise value" of adding a new product to the portfolio and E_c is the entry value for firms city c (see the proof for a complete characterization). In addition, the second stage value function of an entrant firm who is located in city c is

$$V_c^e(\tilde{I}_c, Z_c, A) = \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)Q\right\}.$$

Intuitively, F can be interpreted as the quality-adjusted franchise value of adding a new product to the firm's portfolio, while E_cQ is the value at entry for all firms in city c. Note that E_c does not depend on the firm's portfolio of products, so it does not vary across firms in the same city. The term $\max\{0, E_cQ\}$ in the firm's value function reflects the fact that each firm has the choice to invest in R&D or not, and as such can be interpreted as the option value of investments in innovation. When investing in R&D is optimal, $E_c \geq 0$ and the arrival rate $x_{f,c}$ is strictly positive. However, firms can also choose not to invest in R&D in any given period – for example, if the realized value of the shock Z_c is too low. In this case, the firm does not hire inventors ($i_{f,c} = 0$), produces no innovation ($x_{f,c} = 0$), and does not need to pay the fixed cost ($w_c^i\kappa$).

The assumption that $g < r$ is a technical requirement for the present discounted value of firms to be finite. The intuition is as follows: since the firm's value from investing in R&D grows at rate g (so long as $E_c \geq 0$), if $g > r$, the optimal strategy for any given firm will be to invest as much as possible on R&D (for example, by borrowing capital at rate r). This will generate an expected growth in the firm's value that is larger than its discount factor, and thus, its present discounted value diverges. This issue is avoided when the growth rate is smaller than the real interest rate.

Free Entry. The first stage of the entrant's problem specifies that entrants are free to locate in any city they wish. Since there is a large mass of potential entrants to every city in the economy, free entry implies that, in equilibrium the value of entry must be zero for all cities and at all times:

$$V_c^e(\tilde{I}_c, Z_c, A) = 0 \text{ for all } c \in \{1, \ldots, C\} \text{ and all } t.$$

The intuition behind this condition simple: if the entry value were positive, firms would keep entering the city, which increases congestion and eventually pushes the value of entry to zero. If the entry value were negative (i.e., $E_c < 0$), two things happen. First, no firms will enter the city. Second, incumbents will refrain from investing in R&D (see the discussion above). This pushes down the demand for inventors in the city, reducing congestion costs and the value of entry to zero.

One alternative way to read the free entry condition is that $V_c^e(\tilde{I}_c, Z_c, A) = 0$ for all cities and for *any* value of the shock Z_c and the aggregate state A. In other words, the population of inventors must adjust so that the value of entry is zero in all cities. From proposition 1, this implies that $E_c = 0$ regardless of the value of the state variables. Under this interpretation of the free entry condition, proposition 2 determines the population of inventors in each city (proof in appendix B).

Proposition 2. *Imposing (1) free entry, (2) labor market clearing for both inventors and production workers, and (3) assuming a large number of cities $C \to \infty$, the population of inventors in each city is given by*

$$I_c = I \times \frac{\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}}{\sum_{c=1}^{C} \left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}} \times \frac{Z_c^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}} \tag{6}$$

where $\Theta = (1-\beta)\theta - \psi\eta(1-\theta)$. Moreover, the arrival rate of an innovation for a firm f located in city c is given by

$$x_{f,c} = \left(\kappa \frac{\psi}{1-\psi}\right)^{\psi} \bar{\chi}_c \tilde{I}_c^{\psi\eta} Z_c \tag{7}$$

if the firm chooses to invest in R&D, and zero otherwise. The number of inventors hired by each firm in city c is $i_{f,c} = \frac{\psi}{1-\psi}\kappa$ in case of positive investment and zero otherwise. The number of firms located in city c who invest in R&D in each period is

$$N_c = \left(\frac{1-\psi}{\kappa}\right) I_c. \tag{8}$$

Finally, it is also shown that the population of production workers in city 0 is proportional to L (so L_0 does not vary over time), and that w^i is not affected by city-specific productivity shocks

$$\frac{w^i}{Q} \propto \frac{\bar{\pi} L_0}{r + D} \left\{ \frac{1}{C} \sum_{c=1}^{C} \left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}} \right\}^{\frac{\Theta}{1-\theta}}. \tag{9}$$

Proposition 2 has three important results. First, it shows that the population of inventors in each city has a closed-form solution in equation (6), with relatively simple terms. Cities with higher amenities, higher productivity for innovation, and higher R&D subsidies will have more inventors (assuming $\Theta > 0$; see section 4). The population of inventors in each city also reacts to the productivity shock Z_c, increasing in periods where the shock is larger. The composite parameter Θ can be interpreted as the "net elasticity" of congestion: $(1-\beta)$ captures the elasticity of congestion with respect to a city's population, while $\psi\eta$ is the elasticity of the production of innovation with respect to the population of inventors. Θ is the weighted difference between these two elasticities, where the share of expenditure determines the weights on the tradable good, θ.

Second, proposition 2 says that the *baseline* wage of inventors – and therefore of production workers as well (see equation B.4) – does not react to any of the city-specific productivity shocks. Put differently, even though each city is subject to an idiosyncratic productivity shock, these shocks "average out" on the aggregate, so the economy can still operate on a Balanced Growth Path where there is no aggregate uncertainty.

Third, the proposition shows that the optimal arrival rate of innovation $x_{f,c}$ is uniform across firms located in the same city and who make positive investments in R&D. Note, however, that the expected value of investing in R&D is null because of the free entry condition. As a result, it is possible that some firms in the city choose to invest in R&D and some choose not to. Indeed, the number of inventors hired by firms who make positive investments does not respond to productivity shocks. Instead, the adjustment to those shocks is done entirely on the extensive margin – both by incumbents who decide whether or not to hire inventors and by the entry and exit of firms.

3.4 Determining the Growth Rate

To determine the growth rate of the economy and complete the characterization of the model, we must first compute that aggregate rate of creative destruction, D. Because all firms located in the same city make the same investment in innovation, $D = \sum_{c=1}^{C} N_c x_{f,c}$. Corollary 1 expresses this rate as a function of the spatial distribution of the population of inventors in the economy.

Corollary 1. *Define $\bar{I}_c$ to be the expected value of I_c with respect to the local productivity shocks. Similarly, $\tilde{\bar{I}}_c$ is the expected density of inventors in city c. The aggregate rate of creative destruction is then*

$$D \propto \frac{1}{C} \sum_{c=1}^{C} \bar{\chi}_c \tilde{\bar{I}}_c^{1+\psi\eta}, \tag{10}$$

Corollary 1 shows one of the key takeaways from this model: the aggregate rate of innovation in the economy depends not only on the amount of inventors in the country, but also on how inventors are geographically distributed. Since local R&D subsidies can change the spatial distribution of inventors by attracting more firms to a given location, it follows that changes in the spatial distribution of those subsidies affect the aggregate rate of innovation in the economy – even if the average subsidy rate (or total expenditure) is kept constant. The corollary also proves, as alluded to before, that the rate of creative destruction is indeed constant on the SBGP equilibrium (as both $\bar{\chi}_c$ and $\tilde{\bar{I}}_c$ are fixed over time).

Proposition 3 determines the value of the rate of growth of the economy in the SBGP equilibrium, defined by $g = \dot{Y}/Y$.

Proposition 3. *In a Stationary Balanced Growth Path equilibrium where the production of the final good in the economy grows at rate g, the following are true:*

(1) The average quality of intermediate goods, Q, and the baseline wage for both types of workers, w^i and w^ℓ, all grow at rate g. The utility level of workers, u^i and u^ℓ, grows at rate $(1-\theta)g$.

(2) The value of the rate of growth is $g = \lambda D$.

(3) Let $\mathcal{J}_c(t)$ be the set of intermediate goods that are produced in city c at time t, $Q_c(t) = \int_{\mathcal{J}_c(t)} q_j(t)dj$ their aggregate quality, and $g_c(t) = \dot{Q}_c(t)/Q_c(t)$ the rate of growth of Q_c in time t. For large values of t, we have that

$$\frac{\mathbb{E}[\dot{Q}_c(t)]}{\mathbb{E}[Q_c(t)]} = g \quad \text{or, equivalently,} \quad \mathbb{E}[g_c(t)] = g - \frac{\text{Cov}(g_c(t), Q_c(t))}{\mathbb{E}[Q_c(t)]}.$$

Part (1) of proposition 3 describes the growth rate of aggregate variables in the SBGP equilibrium of the model. Part (2) then shows that the rate of growth of the economy is equal to the rate of creative destruction multiplied by the innovation step-size. Part (3) shows that the ratio between the expected variation in the aggregate quality of goods produced in city c and the expected quality of these goods converges to the national growth rate with time. Intuitively, this result is a consequence of innovation by creative destruction: cities that produce goods with higher-than-average quality will innovate over ("steal") products whose qualities are, on average, lower than the goods already produced in the city, and vice-versa. Through this mechanism, creative destruction acts as a mean reverting force for the quality and the value of goods produced in each city, precluding all economic activity from concentrating in a single city.[13] Lastly, note that none of the results in the proposition are predicated upon the initial spatial distribution of economic activity: as long as no city is large enough to drive the evolution of aggregate variables by itself, proposition 3 holds regardless of where the economy starts from.

4 Identification and Estimation of the Model

The identification and subsequent estimation of the parameters in the model proceeds in three steps. In the first step (section 4.1), we calibrate a set of parameters that can be directly matched to quantities in data or found in similar studies in the literature. The second step (section 4.2) uses linear regressions to estimate the elasticity of the agglomeration spillover with respect to the population of inventors, as well as the elasticity of congestion with respect to the population of production workers in each city. Finally, the third step (section 4.3) identifies the remaining parameters by matching moments in the model to moments in the data. This step-by-step structure simplifies the identification of the model by allowing each step of the process to take as given the values of parameters identified in previous steps.

[13] Another way to see this result is to note that the covariance between the rate of growth of a city and the quality of goods produced there is proportional to $g - \mathbb{E}[g_c]$. Hence, if a city's expected growth rate is larger than g, this city tends to grow faster when the quality of goods produced there is lower.

We use data on patent filings and economic activity within cities to estimate the parameters in the model. Data on patents is available on the USPTO Patent Dataset, which records all patents registered in the US. It also provides information on the identity and location of each patent's inventors and assignees (owners). We use the County Business Patterns Dataset (CBP), which provides information on the demography and economic activity in each county in the US. To approximate the prices of non-tradable goods we use the Zillow Rent Index (ZRI), which estimates the median rental value per square foot across counties.[14] Both the CBP and the ZRI are aggregated to the city level, whose empirical counterpart is a CBSA (core-based statistical areas). Those are geographic areas defined by the US Office of Management and Budget that consist of one or more counties (or equivalents) anchored by an urban center of at least 10,000 people plus adjacent counties that are socioeconomically tied to the urban center by commuting. We use the most recent definitions of CBSAs, based on the 2010 Census standards. Appendix C.1 describes in detail the construction of the dataset. For most of the estimation procedure, we focus on a panel of firms (and their respective locations) who have filed patents between 1998 and 2016.

4.1 Calibration

A subset of the parameters in the model, summarized in table A.1, can be matched to the data in a straightforward way. We start by setting the rate of growth of the economy to $g = 2\%$, which corresponds to the annualized historic rate of growth in the US. The discount rate of consumers is also set to $\rho = 2\%$, and the real interest rate is fixed at $r = 3.8\%$ (this corresponds to the average real interest rate in the US between 1961 and 2017 according to the World Bank).

Innovation. We rely on previous work to determine the value of two parameters in the innovation process: the curvature of the innovation production function, ψ, and the innovation step size, λ. We set $\psi = 0.5$ following several studies that find similar values for this parameter.[15]

The innovation step size is $\lambda = 0.132$. This is the value estimated by Acemoglu et al. (2018) in a setting close to the one presented here. Akcigit and Kerr (2018) find a similar value for the step size of "external" innovations (i.e., innovations that do not target a firm's own products). In both cases, the step size estimation is achieved through a simulated method of moments procedure that targets, among others, firms' sales and R&D costs. Intuitively, these data identify the innovation step size because the increase in the quality of goods after an innovation is reflected in the sales to R&D cost ratio of firms.

[14]One drawback of this series is that it is only available after 2010. As an alternative, we also employ the Zillow Home Value Index (ZHVI) to measure the price of non-tradable goods. The ZHVI has the median house price per square foot in each county in the US, starting in 1996.

[15]A first group of papers identifies ψ as the elasticity of patents with respect to R&D expenditure (Blundell et al., 2002; Griliches, 1990). Acemoglu et al. (2018) arrive at the same value when computing the elasticity of R&D expenditures with respect to scientists' wages using firm-level data from the Census Bureau. Yet another group of researchers identifies this parameter using the elasticity of R&D expenditure with respect to taxes/subsidies, which in this model is equal to $\psi/(1 - \psi)$. In a survey, Hall and Van Reenen (2000) conclude that this elasticity is around unity. The same value is found by Bloom et al. (2002), Wilson (2009), and a recent study by Akcigit et al. (2020). Note that a unit elasticity of patents relative to taxes also implies $\psi = 0.5$.

R&D Subsidies. R&D subsidies are set to the value of existing R&D tax credits in each state (the sum of federal and state-specific credits). Because of differences in state tax codes, the *statutory* credit rates can be different from the *effective* credit rates (i.e., the actual rates that are applied to firms' costs). For example, R&D tax credits only apply to R&D investments over and above a given threshold, which can differ across states. In addition, some states consider the federal tax credit as taxable income, thus "recapturing" part of the credit. We therefore assign s_c to be equal to effective R&D tax credit rate that applies to the highest tier of R&D investments in each state, available through the replication files for Wilson (2009).

Given the model's structure, two additional adjustments were made to the data. First, the value of R&D credits can change over time, while the model assumes that s_c is fixed. The bulk of the variation in tax credits happened before 1995, but there are still some cases where the credit rate changes after that. We use the average credit rate between 1998 and 2006 (the last year available in the data) to measure the R&D subsidy in each state. Second, the model is written in terms of cities (CBSAs), which is some instances do not fall into the geographical boundaries of states. In those cases, we match the CBSA to the state where its largest urban center is located and assume that the R&D credit rate of that state applies to the entire CBSA. For example, the New York-Newark-Jersey City, NY-NJ-PA MSA is matched to the state of New York, the Boston-Cambridge-Newton, MA-NH MSA is matched to Massachusetts, and so on.

Production and Preferences. The elasticity of quality in the production of the final good, ε, coincides with the profit/sales ratio for intermediate good producers, which can be observed in the National Income and Product Accounts tables published by the Bureau of Economic Analysis (BEA). In addition, the preference parameter θ is the share of expenditure on non-tradable goods by consumers. We set this parameter to 0.6, which is roughly the share of expenditure in housing and transportation found in the Consumer Expenditure Survey, published by the Bureau of Labor Statistics (BLS), which also coincides with the share of investment on non-tradable goods found by Bems (2008).

Population. The population of inventors in the economy, I, can be found through the unique inventor ID in the USPTO patent data. For each year, we compute the total number of inventors in the US that are authors of a patent, I^{pat}. Since the number of inventors hired to produce innovation (not counting those that cover the fixed cost) is $\sum_c N_c i_{f,c} = \psi I$, we have that $I = \mathbb{E}[I^{pat}/\psi]$, where the expectation is taken over time.

The population of production workers ("non-inventors") in the economy is set so that $L+I$ matches the total employed population in the CBP data. Finally, L_0 is the total employed population in CBSAs that did not file any patents throughout the sample (again averaged over time). Since the size of the economy does not affect the results in the quantitative exercises, we normalize the population so that the total number of inventors in the economy is 1.

The number of cities is also chosen to match the data. There are 917 CBSAs in the US (not counting Puerto Rico because it is not included in the CBP), of which 860 have filed at least one patent between 1998 and 2016. Given that the number of patents filed measures innovation, we assume that the remaining CBSAs have not produced any innovation over the sample. As a result, $C = 860$ (the cities who have a positive expected productivity in innovation) and city 0 is representative of the remaining 57 CBSAs.

4.2 Linear Regressions

In this section, we estimate η and β, which help to determine the elasticities of agglomeration and congestion, respectively. In both cases, these parameters can be estimated using linear regressions based on relationships predicted by the model.

4.2.1 The Elasticity of Agglomeration

The production function for innovation, equation (1), leads to a log-linear relationship between the innovation arrival rate, the number of inventors hired, and the population of inventors in each city, suggesting that those parameters can be estimated by linear regression. However, the model is written in continuous time, while the data is only available at a yearly rate. Thus, equation (1) must first be transformed to reflect the same frequency as the data. The details for doing so are described in appendix C.2. leading to the following regression

$$\log(x_{f,c,t}) = \psi \log(i_{f,c,t}) + \psi\eta \log(I_{c,t}) + \delta_c + z_{f,c,t}, \tag{11}$$

where, $x_{f,c,t}$ is the number of innovations produced by firm f, located in city c, during year t; $i_{f,c,t}$ is the number of inventors hired by firm f during year t; $I_{c,t}$ is the population of inventors in city c during year t; δ_c is a city fixed-effect; and $z_{f,c,t}$ is a function of the city-specific productivity shocks.

To proxy for the production of innovation in each year, we use the number of patents filed by a firm in that same year. This is a fairly common practice, but it does have some caveats. First of all, not all innovations are patented. Possible reasons for that include firms who decide to protect their intellectual property by other means (for example with trade secrets) or the fact that some innovations are not "patentable" (e.g., new managerial practices or marketing strategies). Second, not all patents represent an innovation over a product. Examples include defensive patenting and patent trolls.

To reduce the potential for a mismatch between patents filed and the production of innovation by firms, we include an additional set of controls into the regression above. First, we include total number of citations that the patents filed by each firm receive in a given year. Patents that do not generate an innovation are less likely to be cited by future patents. Thus, including the number of citations as a control helps to separate innovative patents from other types of patents. In addition, more recent patents mechanically receive fewer citations. To minimize this issue, we interact the number of citations with a dummy for the year in which the patent applications were filed.Second, we add a set of dummies for each sector, which control for the possibility that some sectors are more prone to patenting than others. Finally, we include year fixed effects into the regression to capture aggregate variations over time (for example population growth).

Endogeneity One important prediction of the model is that both $i_{f,c,t}$ ad $I_{c,t}$ are correlated with the local shock $z_{f,c,t}$, so that estimating the coefficients in regression (11) via OLS would recover neither ψ or η. We address this issues in two steps. First, note that the regression can be rearranged to

$$\log(\text{patents}_{f,c,t}) - \psi \log(i_{f,c,t}) = \psi\eta \log(I_{c,t}) + X'_{f,c,t}\Gamma + \delta_c + \delta_t + z_{f,c,t}$$

where $X_{f,c,t}$ includes the controls mentioned above, and δ_t is a year fixed effect. Since the value of ψ is known, the left-hand side of the equation can be constructed in the data. Aggregating to the city level, we find

$$\frac{1}{N_{c,t}} \sum_{f=1}^{N_{c,t}} \left[\log(\text{patents}_{f,c,t}) - \psi \log(i_{f,c,t}) \right] = \psi\eta \log(I_{c,t}) + X'_{c,t}\Gamma + \delta_c + \delta_t + z_{c,t} \tag{12}$$

where $N_{c,t}$ be the number of firms in city c during year t and $z_{c,t}$ is the average shock inside each city. The control vector $X_{c,t}$ includes the average number of citations received by patents filed by firms in city c (interacted with a year dummy) and the employment shares in each NAICS 2-digit sector in city c. This representation simplifies the analysis by removing one of the endogenous variables from the RHS of the equation.

To account for the endogeneity of the population of inventors $I_{c,t}$, we adopt and IV approach that leverages exogenous shocks to industries. We partition the set of products J into K industries, so that each product j can be assigned to a single industry k. The population of inventors in city c can then be written as

$$I_{c,t} = \sum_{k=1}^{K} I_{k,c,t} = \sum_{k=1}^{K} I_{k,c,t-l}(1 + \gamma_{k,c,t-l\to t})$$

where $I_{k,c,t}$ is the number of inventors in industry k living at city c during time t and $\gamma_{k,c,t-l\to t}$ is the rate of growth of $I_{k,c,t}$ between periods $t-l$ and t. Based on this identity, we construct the following instrument for $I_{c,t}$:

$$\mathcal{I}_{c,t,l} = \sum_{k=1}^{K} I_{k,c,t-l}(1 + \gamma_{k,t-l\to t}). \tag{13}$$

where $\gamma_{k,t-l\to t}$ is the overall growth rate of employment in industry k from year $t-l$ to year t. To avoid picking up variation in the total number of inventors in each year, the growth rate is computed using the shares of employment in each industry relative to the total population of inventors. If $I_{k,t}$ is the number of inventors in industry k during year t, $\gamma_{k,t-l\to t} = \frac{I_{k,t}/I_t - I_{k,t-l}/I_{t-l}}{I_{k,t-l}/I_{t-l}}$.

To ensure the exogeneity of the instrument, we follow Autor et al. (2013), and compute the industry growth rates $\gamma_{k,t-l\to t}$ using inventors residing outside of the US who have registered patents with the USPTO (who are responsible for about 50% of all registered patents in the sample). Industries are defined based on NBER's patent subcategories (which add up to 38 different industries) and inventors are assigned to an industry based on the modal sub-class of the patents he or she filed.[16]

The instrument $\mathcal{I}_{c,t,l}$ has a structure that resembles the commonly used "shift-share" research design (Adão et al., 2019; Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020), with the obvious difference that $I_{c,k,t-l}$ is the population level, not a share (see appendix C.3.1 for more details). The economic content is preserved, however: the growth rate of employment in each industry acts as an exogenous shock or "shifter" and the population level $I_{k,c,t-l}$ measures the city's exposure to industry shocks. However, the shift-share structure can also affect the cross-region correlation of the regression residuals (Adão et al., 2019). Intuitively, regions that have a similar industry composition in their population of inventors will also have a similar exposure to the shifters $\gamma_{k,t-l\to t}$, and therefore will tend to have similar residuals as well. As a result, we include two sets of standard errors when the IV strategy is used: one that is clustered across cities and one that is adjusted using the method developed by Adão et al. (2019).

The estimateed coefficients are shown in table 1. The top panel shows the first stage of the IV estimation, and the bottom panel shows the second stage. Column (1) has the OLS estimates. Columns (2) - (4) show the IV estimates with lags l between 5 and 10 years (specified at the bottom of the table). Column (5) has the IV estimates when $I_{i,c,t-l} = I_{i,c,t_{90-95}}$ is fixed at its average level between 1990 and 1995 and the growth rate $\gamma_{k,t_{90-95}\to t}$ is computed using average industry shares in the same period as the base value (recall that the first year in the estimation sample is 1998). All regressions are weighted by the number of firms in each city to account for the fact that the data consists of averages over firms. Standard errors are clustered at the city level to allow for serial correlation of shocks within each city. In the second stage, we present a second set of standard errors based on Adão et al. (2019), which adjust for the for the shift-share structure of the instrument and are indicated by AKM SE. By and large, the estimated coefficients are statistically significant and vary between 0.07 and 0.10.

If compared to other estimates of the elasticity of agglomeration (which usually do not focus on innovation), the values in table 1 are quite large. In fact, Duranton and Puga (2014) state that most studies have found this elasticity to be between 0.02 and 0.05.Innovation can, however, be more responsive to agglomeration spillovers than the production of goods. For example, Carlino et al. (2007) compute the elasticity between the number of patents per capita and employment density in metropolitan/urban areas in the US. In their baseline specification, they find this elasticity to be approximately 0.19. The main difference between that study and this one is that the amount of patents per capita does not account for differences in firm size that arise in different cities – and cities with higher agglomeration will also have more and larger firms, so they naturally produce more patents.

Identification Conditions and Robustness Checks. There are two ways to satisfy the orthogonality condition for the instrument. The first one requires that exposures $I_{k,c,t-l}$ are uncorrelated with the local shock $z_{c,t}$ (Goldsmith-Pinkham et al., 2020). This is unlikely to be true for small values of the lag l if shocks are serially correlated and local shocks differentially affect industry employment shares due to variation in cities' industrial composition. However, this condition becomes plausible when lags are larger, or when the industry employment levels

[16]Note that this means that *industries* are different from *sectors*, as the latter are defined based on the distribution of economic activity in each CBSA according to NAICS; see the data appendix C.1 for more details.

Table 1: Estimation of the elasticity of agglomeration, $\psi\eta$.

First Stage

	(1)	(2)	(3)	(4)	(5)
$\log(\mathcal{I}_{c,t,l})$		0.546^{***}	0.481^{***}	0.398^{***}	0.244^{***}
		(0.031)	(0.033)	(0.035)	(0.042)
F-stat. excluded inst.		315.10	214.79	130.93	33.94

Second Stage

	(1)	(2)	(3)	(4)	(5)
log(Inventors in City)	0.070^{***}	0.104^{***}	0.098^{***}	0.105^{***}	0.104^{*}
	(0.014)	(0.019)	(0.020)	(0.023)	(0.060)
AKM SE		(0.014)	(0.017)	(0.016)	(0.001)
Method	OLS	IV ($l = 5$)	IV ($l = 7$)	IV ($l = 10$)	IV ($l = t - t_{90-95}$)
Observations	11279	11231	11220	11201	11210
Implied η	0.140	0.208	0.196	0.210	0.208

Standard errors are clustered at the CBSA level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively. AKM SE indicates alternative standard errors, calculated according to Adão et al. (2019). All specifications control for patent quality and city industry composition, as well as CBSA and year fixed effects.

are fixed at a period that predates the estimation sample (columns 4 and 5).

Second, the exclusion restriction also holds if the industry growth rates $\gamma_{k,t-l\to t}$ are asymptotically uncorrelated with the industry-specific average of local shocks (Borusyak et al., 2022). Measuring growth rates $\gamma_{k,t-l\to t}$ outside of the US addresses many of the issues that could be raised about the plausibility of this assumption. One concern that remains is that some industries might be highly concentrated in one single city – enough that the city's local shocks are able to affect global trends in that industry (Silicon Valley may come to mind). To address this issue, we re-run the regressions in table 1 after excluding industries for which the employment share in a single city exceeds 15% at any point in time. Varying this threshold between 10 and 25% produces comparable results (see appendix C.3.2).

Another concerns the log-log specification of the regression, which discards observations in which firms have not produced a patent. Since innovation is stochastic, this specification could introduce bias by selecting firms located in cities that are large, or experienced mostly positive shocks. To address this issue, we rewrite (12) as a Poisson regression model, which allows firms to produce no patents in any given year. Appendix C.3.3 describes this regression model in detail and shows the estimated coefficients. The resulting elasticity of agglomeration becomes even higher than above, estimated at approximately $0.13 - 0.15$.

One last robustness check, also described in appendix C.3.3, tests the hypothesis inventors/firms might benefit from other sources agglomeration. For example, firms could benefit from being close to other firms that they can observe and learn from; alternatively, inventors could benefit simply from living in densely populated areas, not necessarily by other inventors. The results we find suggest otherwise: after accounting for the population of inventors, the coefficients on the number of firms, total employment, and total establishments in each city are either negative or not statistically significant.

The value of η. Givent the results in table 1, as well as the robustness checks in appendix C.3, we use $\eta = 0.20$ as the baseline value to compute the optimal distribution of R&D subsidies in section 5. In appendix F, we perform sensitivity analyses using $\eta = 0.15$ and $\eta = 0.25$, which roughly spans the range of estimated coefficients found in all specifications.

We also estimate equation 12 using different values of ψ – namely, $\psi = 0.4$ and $\psi = 0.6$. When $\psi = 0.4$, the estimated coefficient $\psi\eta$ revolves around 0.11 (0.09 if estimated via OLS), which implies $\eta = 0.275$. Conversely, when $\psi = 0.6$, the estimated coefficient $\psi\eta$ fluctuates around 0.09 (0.05 if estimated via OLS), which implies $\eta = 0.15$. In all of the IV specifications, coefficients are statistically significant at the usual levels, and the implied value for η falls in the range used for sensitivity analyses in appendix F.

4.2.2 The Elasticity of Congestion

Once the share of consumer expenditures on non-tradable goods is known, the parameter that determines the elasticity of congestion in the model is the return to scale on the production of non-tradables, β. With constant returns to scale ($\beta = 1$) there is no congestion force, as the production of the non-tradable good scales up with the city size. As β becomes closer to zero, congestion costs become more and more relevant – to the point where the supply of the non-tradable good is fixed and all variation in city size is absorbed into prices. This intuition offers some insight into how best to identify β. The first-order condition of the non-tradable good producer's problem gives $p_{n,c} \propto w_c^{\ell} \left(\frac{L_c}{\bar{m}_c}\right)^{1-\beta}$. Using equation (4) to substitute for wages, the model implies the following empirical relationship (details in appendix C.4)

$$\log(p^h_{c,t}) = \left(\frac{1-\beta}{1-\theta}\right) \log(L_{c,t}) + \delta_c + \delta_t + z^h_{c,t} \tag{14}$$

where $p^h_{c,t}$ is the median rental value per square foot of housing units in city c during year t (a measure of the price on non-tradables), $L_{c,t}$ is the population of non-inventors in city c during year t, δ_c is a city fixed effect that accounts for variation in amenities and land availability, δ_t is a year fixed effect that accounts for the growth in wages/prices and $z^h_{c,t}$ is a city-specific shock.

Given θ, the coefficient on $L_{c,t}$ identified β. However, like the population of inventors in each city, the model also predicts that the population of production workers is correlated with $z^h_{c,t}$. As a result, estimating this regression via OLS will not recover the true value of β. Notwithstanding, given that the population of production workers and the population of inventors in each city is highly correlated, $\mathcal{I}_{c,t,l}$ also serves as an instrument for the population of production workers.

Table 2 displays the estimation results from (14) using $\mathcal{I}_{c,t,l}$ as an instrument for $L_{c,t}$. In this case, the first stage shows a lower predictive value, although the F-statistic generally remains above commonly used thresholds. The implied value of β is consistent across specifications, between 0.5 and 0.6. This is quite large. For comparison, Behrens et al. (2014) find this elasticity to be between 0.08 and 0.09. This difference is due to the inclusion of city fixed effects in my model.[17] Those fixed effects reflect in part the fact that cities have different amenities,

[17]Running this regression in the same data, but without including the city fixed effects, produces coefficients in the range of 0.074 to 0.078, depending on the lag of the instrument.

Table 2: Estimation of the elasticity of congestion, $(1-\beta)/(1-\theta)$.

First Stage

	(1)	(2)	(3)	(4)
$\log(\mathcal{I}_{c,t,l})$	0.020***	0.023***	0.018***	0.023***
	(0.005)	(0.008)	(0.005)	(0.008)
F-stat. excluded inst.	14.58	7.47	12.95	8.89

Second Stage

	(1)	(2)	(3)	(4)
log(Prod. Workers in City)	0.981***	1.013***	1.267***	0.902***
	(0.326)	(0.394)	(0.392)	(0.337)
AKM SE	(0.026)	(0.023)	(0.033)	(0.013)
Method	IV ($l=5$)	IV ($l=7$)	IV ($l=10$)	IV ($l=t-t_{90-95}$)
Observations	2855	2849	2845	2846
Implied β	0.608	0.595	0.493	0.639

Standard errors are clustered at the CBSA level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively. AKM SE indicates alternative standard errors, calculated according to Adão et al. (2019). All specifications control for CBSA and year fixed effects.

which affect the utility of consumers. Leaving them out of the regression can therefore severely bias the elasticity of prices with respect to the population, since individuals are willing to pay higher prices to live in cities where amenities are higher.

Identification Conditions and Robustness Checks. The structural error in equation (14) is a function of the same productivity shock that determines the residual in the previous section. Hence, the conditions for the orthogonality of the instrument in both cases are also the same. Table 2 shows the estimated elasticity of congestion when the lag l used to construct the instrument varies from 5 to 10 years and when industry employment shares in each city are fixed at their average level between 1990 and 1995 – leveraging the fact that for large l, $I_{c,k,t-l}$ and $z^h_{c,t}$ are likely to be uncorrelated. Appendix C.4 presents these same results when using an instrument that excludes industries whose employment share (of inventors) in any single city share exceed 15% at any point in time.

Rental values are only available in the ZRI database after 2010, which explain the small number of observations in table 1. Because of that, we also an alternative version of (14), where $p^h_{c,t}$ is approximated by the median housing *price* in each city (a series that goes back to 1996)[18]. This regression implies a higher value or β, around 0.8, which reflects the fact that housing prices tend to be less elastic than rental prices.

The value of β. Combining the estimation results in table 2 and in appendix C.4, we adopt $\beta = 0.6$ as the baseline value to compute the optimal distribution of R&D subsidies in section 5. We also perform sensitivity analyses using $\beta = 0.5$ and $\beta = 0.8$, which spans the range of

[18]Using housing prices to approximate $p^h_{c,t}$ presents its own issues, as houses can also be thought of assets, whose prices might reflect agents' expectations about the future of the economy.

estimated coefficients found in all specifications.

4.3 Moment Matching

Fixed Cost of Innovation. To estimate size of the fixed cost of innovation, κ, we use equation (8), which relates the number of firms in each city to the number of inventors in the city. Summing both sides of that equation over cities and rearranging gives

$$\kappa = (1-\psi)\frac{I}{N}$$

where N is the total number of firms in the economy. Given that the average number of inventors per firm in the data is $I/N \approx 21.07$ and $\psi = 0.5$, this relationship gives $\kappa = 10.53$.

City-Specific Parameters. Next, we turn to the set of city amenities α_c and mean productivities $\bar{\chi}_c$ in each city. For cities $c \in \{1, \dots, C\}$, these parameters can be identified off the average share of inventors and patents filed by cities over time. Specifically, the average share of inventors in city c over time is[19]

$$(\text{avg. share of inventors})_c = \frac{\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}}{\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}} \equiv \frac{\bar{I}_c}{I}.$$

Similarly, the average share of patents filed in each city is

$$(\text{avg. share of patents filed})_c = \frac{1}{T}\int_0^T \frac{N_c(t)x_{f,c}(t)}{\sum_{c=1}^{C} N_c(t)x_{f,c}(t)}dt = \frac{\bar{\chi}_c \bar{I}_c^{1+\psi\eta}}{\sum_{c=1}^{C}\bar{\chi}_c \bar{I}_c^{1+\psi\eta}}.$$

The two sets of equations above identify α_c and $\bar{\chi}_c$ for every c up to a constant. Since α_c is a preference parameter, its level does not have much meaning and we normalize $\mathbb{E}_c[\alpha_c] = 1$. The scale of $\bar{\chi}_c$ can be identified off equation (10) by imposing that the rate of growth of the model $g = \lambda D$ equals 2%, the historic annualized rate of growth in the US. The effective amenity level in city 0 can be found by matching the share of population in city 0, $L_0/(I+L)$, to its counterpart in the data. Both of these procedures are described in detail in appendix D.1.

Law of Motion of the Productivity Shock Lastly, we describe the identification of shock distribution parameters σ and ϕ. Since only the ratio σ^2/ϕ matters for the equilibrium of the model, we set $\phi = 1$. Next, σ can be found by matching the model-generated cross-sectional variance of the population of inventors between cities with the same moment in the data. Appendix D.2 derives the expression for this variance in the model and shows how to identify σ.

[19] Formally, model's equilibrium gives $\frac{1}{T}\int_0^T \frac{I_c(t)}{I}dt = \frac{\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}}{\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}} \times \frac{1}{T}\int_0^T \frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{\exp\left(\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}\right)}dt$. Since $Z_c(t)$ is an ergodic process, the integral in this expression converges to an expected value when $T \to \infty$ (Bergelson et al., 2012). Assuming that the number of periods available in the data is large enough, the result follows.

4.4 Comparison to Untargeted Moments

We assess the model's external validity by measuring how well it can fit the spatial distribution of variables that were not targeted for estimation. Figure A.3 plots the model's predictions against the data for four untargeted variables: the share of firms per city (panel a), the average number of patents per firm in each city (panel b), the share of total employed population per city (panel c) and the spatial distribution of patents per capita (panel d). In general the model matches those distribution quite well – the correlation between the share of firms per city in the model and data is particularly high, at about 0.97. The distribution of patents per firm is harder to match, as there are many cities – of widely different sizes – that have on average a single patent per firm.

Panels (c) and (d) of figure A.3 show the match between the total employed population and patents per capita between model and data. As a general rule, the model tends to underestimate the total population in cities where there is a small number of inventors and overestimate the population of cities where many inventors live. This happens because the model predicts that the population of inventors and production workers is proportional across cities; in practice, cities tend to specialize in one type of activity. As a result, the model-predicted total population and model-predicted patents per capita tend to be off in each end of the city size distribution. Nevertheless, the correlations between model and data are still fairly high, as shown in panel A of table 3.

Panel B of table 3 compares the outcomes in the model and data at different sections of the city size distribution. Specifically, it ranks cities based on their average population of inventors between 1998 and 2016 and divides them into five bins with an equal number of cities. It then compares the share of firms and the average number of patents per firm in each of those quintiles separately.For reference, we also include share of inventors and patents in each quintile of the city size distribution (there is no comparison between model and data in those cases, as the match is one-to-one).

Table 3: Spatial Distribution of Untargeted Variables

Panel A: Correlations Between Untargeted Variables: Model and Data

Number of Firms	Patents per Firm	Employed Population	Patents per Capita
0.97	0.58	0.81	0.49

Panel B: Spatial Distribution of Untargeted Variables

Bin	Share of Inventors	Share of Patents	Share of Firms (model)	Share of Firms (data)	Avg. Patents/Firm (model)	Avg. Patents/Firm (data)
1	0.002	0.001	0.002	0.003	1.052	0.81
2	0.006	0.002	0.006	0.007	1.006	0.875
3	0.013	0.006	0.013	0.017	1.234	0.993
4	0.04	0.018	0.04	0.045	1.271	1.171
5	0.939	0.973	0.939	0.927	2.181	2.894

4.4.1 Can R&D Tax Credits Shift the Spatial Distribution of the Economy?

The last question addressed in this section is the extent to which R&D tax credits can influence the location of firms. Moretti and Wilson (2017) offer evidence that inventors are very sensitive to state taxes, but R&D tax credits tend to have a smaller effect than other forms of taxation. Similarly, Slattery (2019) finds that state-level subsidies have an important effect over firms' locations, but this effect includes all discretionary state subsidies and cannot be attributed to R&D tax credits alone. Understanding the extent to which R&D tax credits can change the spatial distribution of the economy is relevant for interpreting the counterfactual results in the forthcoming sections, which assess the welfare effects of alternative spatial configurations of those credits.

To answer this question, we leverage the variation of R&D tax credits over time and measure how well the model can predict the spatial dispersion of the economy in the years when the spatial distribution of R&D tax credits differed from what it is today. We focus on the spatial distribution of the population of inventors and of patents filed in the decades of 1970-1979, 1989-1989 and 1990-1999. Using averages across longer periods has two advantages. First, the equilibrium of the model assumes a BGP, so its predictions do not apply to short-term variations. Second, these three decades roughly coincide with broad trends in the adoption of R&D subsidies: in the 1970's, there were no subsidies. In the 1980's, the federal government introduced a (spatially uniform) subsidy; in addition, a few states started offering local subsidies of their own, on top of the federal benefits. By the 1990's, similar policies had been adopted by most states.

We construct the model-implied distribution of inventors and patents per city in any given year by simply providing the model with the value of the R&D tax credit rates in that year (keeping all other parameters fixed). Using those credit rates and the parameters estimated above, we construct the share of inventors and patents produced in each city for all years, then calculate their averages for each decade. Panel A of table A.3 shows the correlations between model outcomes and the data for each decade. For better visualization, we again aggregate cities according to quintiles of the city size distribution (where cities are ordered according to their average population of inventors) and report the model-predicted and observed share of inventors and patents in each of those bins.

To account for persistence in city size, we also analyze the model- and data-implied *changes* in the shares of inventors and patents in each city across time. To this end, we again compute the average share of inventors and patents produced in each city during the 1970's, 1980's, 1990's and 2000's. Next, we compute the difference between those shares in the decades of 1970, 1980 and 1990 relative to their value in 2000. Figure A.4 shows the correlation between model and data outcomes in each decade, which hover around 0.6 for changes in the share of inventors and 0.8 for changes in the share of patents filed. Panel B of table A.3 displays the output of regressing the changes in shares observed in the data on its counterpart in the model, where only the value of the R&D tax credit is allowed to vary.

These results suggest that R&D tax credits are quite relevant for the location decisions of inventors/firms and the production of innovation. The R-squared of the regressions in table A.3 show that changes in the R&D tax credit rate can explain about 40% of the variation of changes in population shares over time and over 60% of the variation of changes in the production of patents. The correlations in Panel A of that same table indicate that those changes often go in the direction predicted by the model.

R&D Tax Credits vs Corporate and Labor Income Taxes. Previous research has linked changes in corporate and labor income taxes to changes in the quantity and location of innovation (Akcigit et al., 2020). As a result, it is important to check whether R&D tax credit rates have moved in tandem with those taxes – which would mean that the results above might not be related to R&D tax credits themselves, but the effects of changes in other local policies.

To test this possibility, we calculate, for each state in the US, the year-over-year change in (1) the statutory R&D tax credit rate, (2) the state corporate income tax rate, and (3) the state labor income tax rate for the top income bracket.[20] Data for all series is available between 1977 and 2006. Across all states, the correlation between changes in the R&D tax credit rate and either corporate or labor income taxes is very small (-0.01 and -0.03, respectively), and not statistically significant at the 10% level. Different specifications of a regression of R&D tax credits on both corporate and labor income taxes all yield the same result: coefficients that are small and not statistically significant (see table A.2).

5 The Welfare Effects of Spatial Policies

Finally, we turn to the main questions motivating this study: can a redistribution of local R&D subsidies increase aggregate welfare in the economy? And if so, what is their optimal distribution? To answer the first question, we compare the current spatial distribution of R&D subsidies in the US with a spatially homogeneous subsidy that is implemented with the same amount or resources.[21] To answer the second question, we assume the existence of a social planner/government that is able to choose the value of all local R&D subsidies in order to maximize welfare. We compute an approximate solution for this problem, which provides a set of optimal R&D subsidies that can be used to measure the potential welfare gains from the redistribution of local R&D subsidies in the US and to inform us about which places should benefit from R&D policy.

5.1 The Government's Problem

Aggregate welfare in this model is measured as the sum of the utility of all workers in the economy, since all firm- and land-owners are fully taxed. For convenience, we assume that the cost of producing $\bar{G}$ units of the public good is $\gamma(\bar{G}) = \bar{\pi}\bar{G}Q(t)$, and that the amount of public good produced is fixed. Let $\Pi(t)$ be the aggregate flow of profits by all firms in the economy in period t. The government's problem is

[20] Data on state income taxes are obtained through NBER's TAXSIM model (`https://users.nber.org/~taxsim/state-rates/`).

[21] In practice, each state is able to choose its own tax credit level, so the spatial distribution of R&D subsidies US can be understood as the outcome of a non-cooperative game played by policy makers in each state. Comparing this decentralized outcome with a spatially neutral subsidy informs us about the welfare gains of allowing states to compete by choosing R&D policy.

$$\max_{\{s_c\}_{c=1}^C} \quad \int_0^\infty e^{-\rho t} \left\{ \sum_{c=0}^C \left[L_c(t) u^\ell(t) + I_c(t) u^i(t) \right] \right\} \bar{G} dt$$

$$\text{s.t.} \quad \int_0^\infty e^{-rt} \left[\sum_{c=1}^C s_c w_c^i(t) I_c(t) + \gamma(\bar{G}) \right] dt = \int_0^\infty e^{-rt} \left[p_{m,0}(t) \bar{m}_0 + \sum_{c=1}^C p_{m,c}(t) \bar{m}_c + \Pi(t) \right] dt.$$

Note that the population of inventors and production workers, their wages, utility levels, land prices, profits, and the rate of creative destruction are all endogenous variables that depend on the full spatial distribution of subsidies. Appendix E shows that this dynamic problem can be transformed into a static one, after plugging in the expression for those variables in the model's equilibrium.

The static version of the government's problem also highlights one important tradeoff that the government has to face. On the one hand, the government's objective function increases with the rate of creative destruction, D: a higher rate of creative destruction means that the economy grows at a higher rate, which therefore implies a higher present value of welfare. Furthermore, from corollary 1, $D \propto \frac{1}{C} \sum_{c=1}^C \bar{\chi}_c \tilde{\bar{I}}_c^{1+\psi\eta}$ which means that a more spatially concentrated population leads to a higher rate of innovation – especially if the population is concentrated in the most productive cities. On the other hand, the normalized congestion-adjusted wage $w^i(t)/Q(t)$ decreases with the rate of creative destruction (see equation 9 in proposition 2). Intuitively, when the rate of creative destruction increases, so does the rate at which firms discount the future, $r + D$, because the probability that any of the firm's product lines will be stolen by a competitor increases. This leads firms to decrease investments in R&D, which reduces the demand for inventors and pushes their wage down (the same happens for production workers through general equilibrium effects). Lower wages, in turn, result in lower welfare.

5.2 A Spatially Homogeneous Subsidy

The current spatial distribution of R&D subsidies in the US can be understood as the outcome of the competition among states to attract innovative firms and inventors into their jurisdiction. To evaluate the effects of this competition, consider a counterfactual economy where states are not allowed to compete, so that R&D subsidies are fixed over space $s_c \equiv \bar{s}$ for all c. Since taxes and other government expenditures are kept constant throughout all counterfactuals, $\bar{s}$ is fully determined by the government's budget constraint. Given the model's calibration, this subsidy rate is close to 19%. For comparison, the average subsidy rate under the current distribution is about 16%.

Moving to a spatially homogeneous subsidy spreads the population of inventors more evenly over space: the HHI index of the city population shares moves from 0.027 to 0.025. Under this alternative population distribution, aggregate welfare falls by 0.77% due to a decrease in the growth rate of the economy of approximately 0.03 percentage points. In contrast, the static baseline wage $\bar{w}^i$ increases by 0.91%. In other words, the decentralized adoption of R&D subsidies by states led to a higher welfare level than what would be attained under a spatially neutral subsidy that is implemented using the same amount of resources. This suggests that the states that offer the largest R&D tax credits are indeed to ones that are comparatively better at producing innovation (leading to a higher growth rate).

5.3 Approximating the Optimal R&D Subsidies

What is the best outcome that we can achieve by just redistributing subsidies across space? Finding the exact optimal subsidies that solve (E.1) can be computationally challenging, as this a non-convex problem with 860 choice variables (cities). Therefore, we compute an approximate solution by imposing a functional form to s_c:

$$s_c = \begin{cases} \zeta \alpha_c^{\xi} \bar{\chi}_c^{\omega}, & \text{if } \zeta \alpha_c^{\xi} \bar{\chi}_c^{\omega} \leq \tau; \\ \tau, & \text{if } \zeta \alpha_c^{\xi} \bar{\chi}_c^{\omega} > \tau. \end{cases}$$

This functional form is motivated by the fact that cities only differ from each other because of either α_c or $\bar{\chi}_c$ – and therefore any differences in the optimal subsidy across cities will necessarily be driven by differences in these two parameters. In addition, we also introduce a subsidy cap, taking values in $\tau \in \{0.3, 0.4, 0.5\}$. This is done to avoid what might be a politically unfeasible subsidy – the highest credit rate currently offered in the data (combining state and federal tax credits) coincides with the lowest value of the cap, at about 30%. In each case, the parameters ξ and ω are chosen in the interval $[-5, 10]$ to maximize aggregate welfare. The scale parameter ζ ensures that the budget constraint is satisfied.

Figure A.5 plots aggregate welfare as a function of ξ and ω (fixing $\tau = 0.4$) and the resulting optimal subsidy as a function of amenities and local productivity. It is clear from panel (b) that the welfare is maximized when innovation is concentrated in cities with high amenities and high productivity, so the optimal subsidy rates should move the economy in this direction. This result is intuitive: cities with high productivity produce more innovation per worker, so moving the population to those cities will generate a higher growth rate. Alternatively, workers living in cities with high amenities will accept relatively lower wages, so firms in those cities experience less congestion costs, all else equal.

Figure 2 maps the optimal R&D tax credits for each city in the US. In accordance to the discussion above, there are two areas that are heavily subsidized under the optimal policy: the Silicon Valley (San Jose) and New York City, which are already the two largest producers of patents in the country. Figure A.6 shows how the optimal R&D subsidies affect the geographical distribution of inventors and patents produced in each city, relative their current values. Note that a big part of the effect of the optimal subsidies is to move the population from mid-sized cities to a few high productivity/high amenity cities, dramatically increasing their share of the population and innovation.

The welfare gains from the spatial reallocation of the population caused by the optimal distribution of R&D subsidies is shown in table 4. When capping the city-level subsidy at 50%, the model predicts that total welfare would grow by at least 6% if the optimal distribution of R&D subsidies was adopted. This gain is generated in part by an increase of 0.26 percentage points in the rate of growth of the economy. However, as mentioned above, baseline wages also fall by over 7%, indicating that the higher rate of creative destruction has lowered the demand for labor by individual innovative firms.

Figure 2: Optimal R&D Tax Credit Rates per City

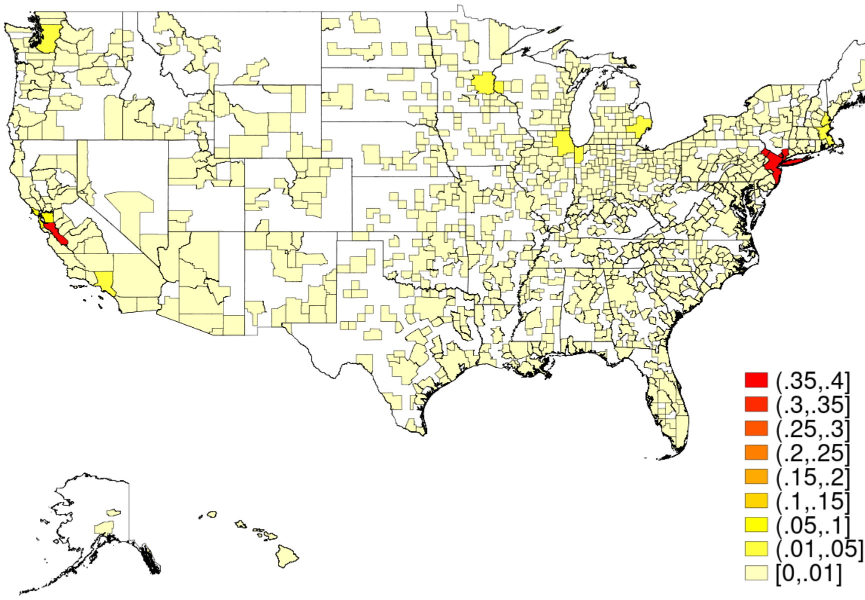


Table 4: Gains from adopting optimal subsidies.

Δ Welfare	Δ Congestion-adj. Wage	Δ Creative Destruction	Δ Rate of Growth	τ
Panel A: City-level Subsidies				
2.95%	−3.60%	0.97 p.p.	0.13 p.p.	0.3
5.23%	−6.38%	1.70 p.p.	0.22 p.p.	0.4
6.15%	−7.75%	2.00 p.p.	0.26 p.p.	0.5
Panel B: State-level Subsidies				
2.50%	−3.65%	0.88 p.p.	0.12 p.p.	0.3
3.06%	−4.36%	1.07 p.p.	0.14 p.p.	0.4
3.23%	−4.71%	1.13 p.p.	0.15 p.p.	0.5

5.3.1 Subsidies by State

The results described above are predicated on the assumption that R&D subsidies can vary by city.Historically, these subsidies were chosen at the state level. Taking the geographical scope of the policy as given, we re-run the exercise above while constraining subsidies to be constant within states. To do that, let $c(\mathcal{S})$ indicate a city c that is located in state $\mathcal{S}$. Denote by $C(\mathcal{S})$ the total number of cities in each state and the approximate optimal subsidy by

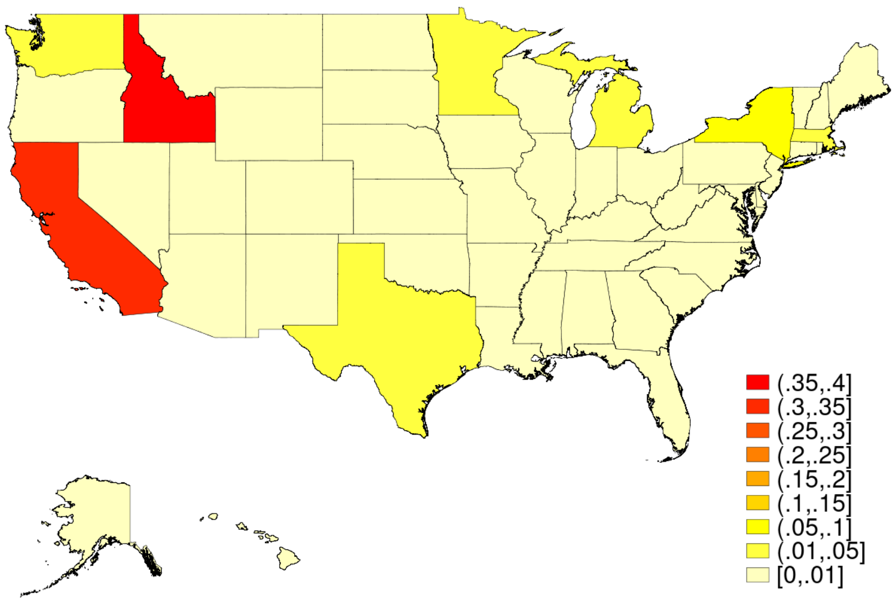
$$s_{c(\mathcal{S})} = \begin{cases} \zeta \frac{1}{C(\mathcal{S})} \sum_{c=1}^{C(\mathcal{S})} \alpha_c^{\xi} \bar{\chi}_c^{\omega}, & \text{if } \zeta \frac{1}{C(\mathcal{S})} \sum_{c=1}^{C(\mathcal{S})} \alpha_c^{\xi} \bar{\chi}_c^{\omega} \leq \tau; \\ \tau, & \text{if } \zeta \frac{1}{C(\mathcal{S})} \sum_{c=1}^{C(\mathcal{S})} \alpha_c^{\xi} \bar{\chi}_c^{\omega} > \tau. \end{cases}$$

Once again, $\tau \in \{0.3, 0.4, 0.5\}$ and the parameters ξ and ω are chosen to maximize total welfare in (E.1). The parameter ζ ensures the the government's budget constraint is satisfied. Panel B of table 4 shows the welfare effects of the optimal distribution of R&D subsidies across states. Note that the distribution follows the same pattern as above, where a higher spatial concentration of innovation leads to gains in welfare. Those gains are smaller, however, as there are more constraints on the value of the subsidy. Finally, note that this pattern is again robust to different values of the agglomeration and congestion elasticities, as shown by the sensitivity analyzes in appendix F.

Figure 3 shows the optimal value of the subsidy in each state. It is interesting to see that the optimal subsidy distribution changes completely, to California and Idaho! The reason for this difference is that the state of New York has a number of mid-sized cities that, when combined, produce a sizable share of innovation in the state. Therefore, subsidizing New York state would bring more inventors to NYC, but also to all other cities, undermining the concentration of the population in a single location.

In contrast, innovation in Idaho is much more concentrated in Boise – which is itself an innovation hub as well. In fact, both NYC and Boise are home to about half of the population in their respective states; however, NYC only produces around 5% of patents in the state of New York, while Boise concentrates almost 60% of the patents in Idaho. As a result, the effect of an R&D subsidy in Idaho would be highly focused in that city, leading to a larger concentration of the population in one single place. One important conclusion from this discussion is that the spatial distribution of the optimal R&D subsidy can drastically change depending on the geographical scope of the policy.

Figure 3: Optimal R&D Tax Credit Rates per State



5.4 Discussion

There are a few important points to keep in mind when interpreting the results found in this section. First, the welfare gains reported here are the product of a pure redistribution of R&D subsidies over space. Expenditures on those subsidies are kept constant throughout all counterfactual exercises, with the exception of endogenous changes in the government's revenue caused by the reallocation of the population. As such, the optimal subsidy rates computed here require no changes in taxation whatsoever.

Second, the gains reported in table 4 are only a lower bound for the increase in welfare that can be obtained by the redistribution of R&D subsidies. This is a consequence of imposing a functional form to approximate the optimal subsidies, which does not necessarily describe the policy that maximizes the government's problem. As a counterpoint, the welfare gains under the optimal distribution of R&D subsidies assumes the existence of a benevolent social planner, which could be in contrast to how those policies were historically adopted. In the US, many of the R&D tax credits were chosen independently by several states, whose goals might be maximizing aggregate welfare and lack coordination across locations (Ossa, 2015).

A third point concerns simplifying assumptions in the model used to compute the optimal policies and aggregate welfare. As discussed in section 2, the introduction of moving costs can have relevant effects on welfare and on the the distribution of the optimal R&D tax credits. Those impacts can take the form of one-time costs paid by workers to move between cities, but could also affect the optimal distribution itself if the costs of moving workers from city A to city B are too large. We argue that the results found here should be thought of as long-run responses to changes in policy, but recognize that the adjustment might not trivial and take some time to complete.

Lastly, placed-based policies often need to balance a tension between subsidizing more productive regions and redistributing resources to poorer regions (Glaeser and Gottlieb, 2008; Henkel et al., 2021). Because the utilities of workers equate across regions, issues related regional inequality cannot be addressed here. As such, we do not argue that *all* policies should work to concentrate resources in a handful of cities. Instead, we focus on R&D tax credits, a policy designed with the clear objective to increase the amount of innovation in the country. Given this goal, my framework is able to speak to both the spatial and dynamic incentives of R&D policies at the firm level and propose a reallocation of resources that would lead to higher growth by concentrating inventors in the most productive cities.[22]

6 Conclusion

This paper assesses whether there are welfare gains from the spatial reallocation of R&D tax credits in the US. As a framework to analyze counterfactual spatial distributions of the tax credit, we construct an endogenous growth model with spatial heterogeneity and agglomeration economies in the production of innovation. This framework contributes to the literature on endogenous growth by nesting a model of growth through creative destruction into a spatial setting. It also contributes to the literature on spatial and dynamic models by developing a tractable model that can be easily matched to micro data.

Qualitatively, we identify an important tradeoff in the the optimal spatial distribution of R&D subsidies: increasing the geographical concentration innovation in highly productive cities will increase the rate of growth of the economy, but it also increases the rate at which firms

[22]Note that this prescription is not inconsistent with alleviating regional inequalities either, as smaller cities can also benefit from higher growth, or focus on other economic activities that are their comparative advantage (see Rossi-Hansberg et al., 2019).

discount the future due to a higher rate of creative destruction. This reduces individual firms' investments in R&D, which puts downward pressure on the wages of inventors and decreases aggregate welfare.

We find that concentrating the population of inventors in cities with high amenities and high productivity has positive and potentially large impacts on aggregate welfare. Furthermore, those gains are achieved through a pure redistribution of the R&D subsidy over space, keeping all taxes and other government expenditures constant. Importantly, the geographical scope of policies matters. While it is optimal to subsidize cities with high productivity/amenities, this does not translate to the state level, as the composition of cities within each state needs to be taken into account.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., and Kerr, W. (2018). Innovation, Reallocation, and Growth. *American Economic Review*, 108(11):3450–91.

Adão, R., Kolesár, M., and Morales, E. (2019). Shift-Share Designs: Theory and Inference. *The Quarterly Journal of Economics*, 134(4):1949–2010.

Aghion, P. and Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, 60(2):323–351.

Akcigit, U., Ates, S., and Impullitti, G. (2018). Innovation and Trade Policy in a Globalized World. NBER Working Paper 24543, National Bureau of Economic Research.

Akcigit, U., Baslandze, S., and Stantcheva, S. (2016). Taxation and the International Mobility of Inventors. *American Economic Review*, 106(10):2930–2981.

Akcigit, U., Grigsby, J., Nicholas, T., and Stantcheva, S. (2020). Taxation and Innovation in the Twentieth Century. *The Quarterly Journal of Economics*, 137(1):329–385.

Akcigit, U. and Kerr, W. (2018). Growth through Heterogeneous Innovations. *Journal of Political Economy*, 126(4):1374–1443.

Austin, B., Glaeser, E., and Summers, L. (2018). Jobs for the Heartland: Place-Based Policies in 21st-Century America. *Brookings Papers on Economic Activity*, 49:151–255.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103(6):2121–2168.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, 22(13):507–553.

Bems, R. (2008). Aggregate Investment Expenditures on Tradable and Nontradable Goods. *Review of Economic Dynamics*, 11(4):852–883.

Bergelson, V., Leibman, A., and Moreira, C. G. (2012). Form discrete- to continuous-time ergodic theorems. *Ergodic Theory and Dynamical Systems*, 32:383–426.

Bloom, N., Griffith, R., and Van Reenen, J. (2002). Do R&D tax credits work? Evidence from a panel of countries 1979–1997. *Journal of Public Economics*, 85(1):1–31.

Blundell, R., Griffith, R., and Windmeijer, F. (2002). Individual Effects and Dynamics in Count Data Models. *Journal of Econometrics*, 108(1):113–131.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *Review of Economic Studies*, 89(1):181–213.

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

Carlino, G. A., Chatterjee, S., and Hunt, R. M. (2007). Urban Density and the Rate of Invention. *Journal of Urban Economics*, 61:389–419.

Correia, S., Guimarães, P., and Zylkin, T. (2019). ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects.

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

Duranton, G. (2007). Urban Evolutions: The Fast, the Slow, and the Still. *American Economic Review*, 97(1):197–221.

Duranton, G. and Puga, D. (2014). The Growth of Cities. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 2, chapter 5, pages 781–853. Elsevier.

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

Fajgelbaum, P. and Gaubert, C. (2020). Optimal Spatial Policies, Geography and Sorting. *Quarterly Journal of Economics*, 135(2):959–1036.

Fajgelbaum, P., Morales, E., Serrato, J. C. S., and Zidar, O. (2019). State Taxes and Spatial Misallocation. *Review of Economic Studies*, 86(1):333–376.

Farrokhi, F. (2021). Skill, Agglomeration, and Inequality in the Spatial Economy. *International Economic Review*, 62(2):671–721.

Gaubert, C. (2018). Firm Sorting and Agglomeration. *American Economic Review*, 108(11):3117–53.

Glaeser, E. L. and Gottlieb, J. D. (2008). The Economics of Place-Making Policies. *Brookings Papers on Economic Activity*, pages 155–239.

Glaeser, E. L. and Hausman, N. (2019). The Spatial Mismatch Between Innovation and Joblessness. NBER Working Papers 25913, National Bureau of Economic Research.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8):2586–2624.

Gompers, P. and Lerner, J. (2001). The Venture Capital Revolution. *Journal of Economic Perspectives*, 15(2):145–168.

Goolsbee, A. (1998). Does Government R&D Policy Mainly Benefit Scientists and Engineers? *American Economic Review*, 88(2):298–302.

Griliches, Z. (1990). Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, 28(4):1661–1707.

Hall, B. and Van Reenen, J. (2000). How Effective are Fiscal Incentives for R&D? A Review of the Evidence. *Research Policy*, 29(4-5):449–469.

Henkel, M., Seidel, T., and Suedekum, J. (2021). Fiscal transfers in the spatial economy. *American Economic Journal: Economic Policy*, 13(4):433–68.

Hsieh, C.-T. and Moretti, E. (2019). Housing Constraints and Spatial Misallocation. *American Economic Journal: Macroeconomics*, 11(2):1–39.

Joint Committee on Taxation (2010). Estimated Budget Effects of the Revenue Provisions Contained in the President's Fiscal Year 2011 Budget Proposal, JCX-7-10R.

Klette, T. J. and Kortum, S. (2004). Innovating Firms and Aggregate Innovation. *Journal of Political Economy*, 12(5):986–1018.

Kline, P. and Moretti, E. (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.

Manso, G. (2011). Motivating Innovation. *The Journal of Finance*, 66(5):1823–1860.

Moretti, E. (2021). The Effect of High-Tech Clusters on the Productivity of Top Inventors. *American Economic Review*, 11(10):3328–2275.

Moretti, E. and Wilson, D. J. (2017). The effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. *American Economic Review*, 107(7):1858–1903.

Ossa, R. (2015). A Quantitative Analysis of Subsidy Competition in the U.S. NBER Working Papers 20975, National Bureau of Economic Research.

Rosenthal, S. S. and Strange, W. C. (2001). The Determinants of Agglomeration. *Journal of Urban Economics*, 50:191–229.

Rossi-Hansberg, E., Sarte, P.-D., and Schwartzman, F. (2019). Cognitive hubs and spatial redistribution. Working Paper 26267, National Bureau of Economic Research.

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard Uiversity Press, Cambridge, MA.

Shiryaev, A. (1996). *Probability*. Springer-Verlag New York, 2 edition. Translated by S. S. Wilson.

Slattery, C. R. (2019). Bidding for Firms: Subsidy Competition in the U.S. Working paper.

Stokey, N. (2008). *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press.

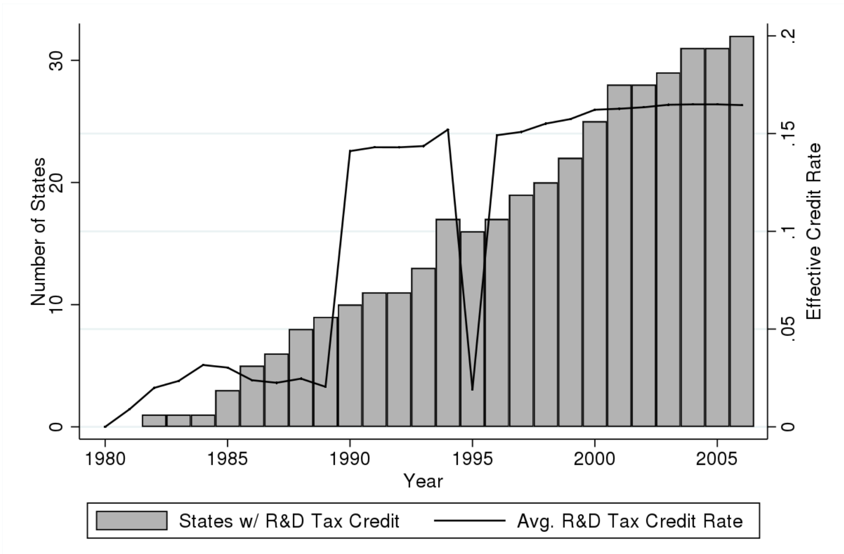
Wilson, D. J. (2009). Beggar Thy Neighbor? The In-State, Out-of-State, and Aggregate Effects of R&D Tax Credits. *The Review of Economics and Statistics*, 91(2):431–436.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA.

Online Appendix for "R&D Subsidies and the Geography of Innovation"

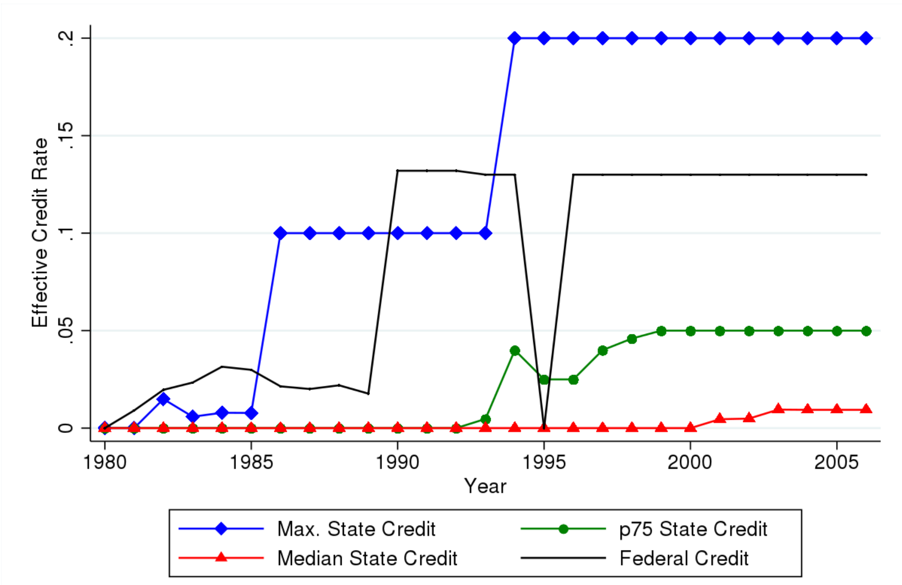
A Additional Figures and Tables

Figure A.1: Number of States that Adopted R&D Tax Credits



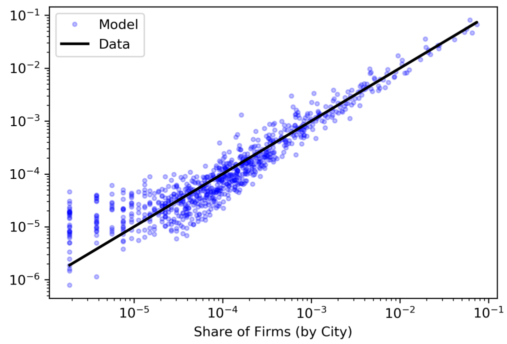
Note: figure shows the average *effective* R&D tax credit rate in the US. See the note in figure 1.

Figure A.2: Distribution of R&D Tax Credit Rates Across States

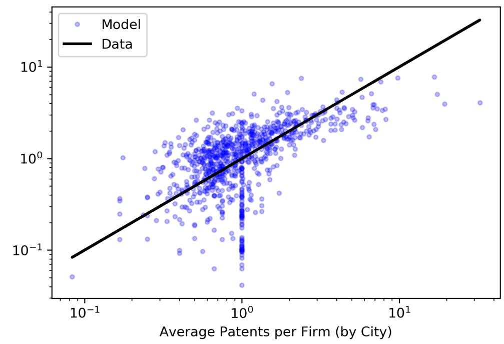


Note: figure shows the *effective* R&D tax credit rates in the US. See the note in figure 1. The two discontinuities in the federal credit rate are due to (1) a change in the method for computing the federal base level in 1991 and (2) the fact that there were no federal credits in 1995.

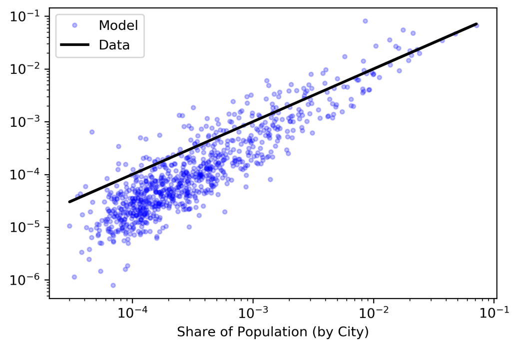
Figure A.3: Model Fit



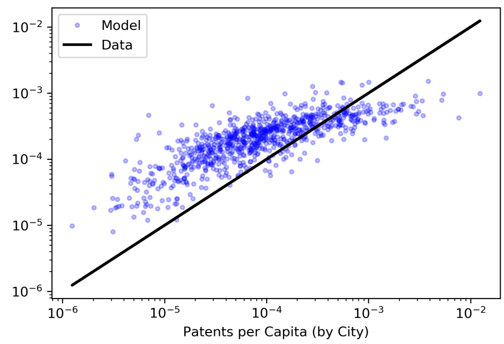
(a) Spatial distribution of firms



(b) Distribution of patents per firm

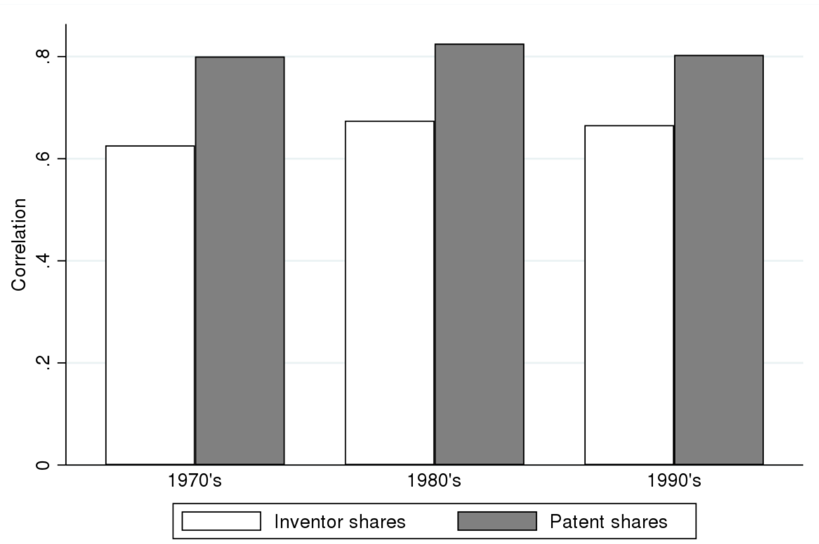


(c) Spatial distribution of employed population



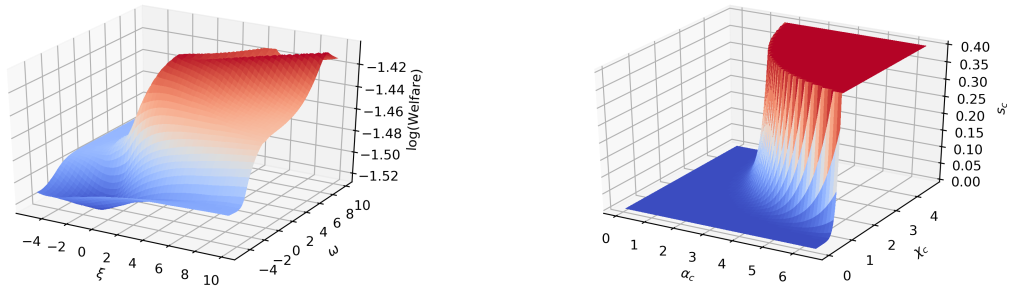
(d) Distribution of patents per capita

Figure A.4: Correlation Between Changes City Population: Model vs Data



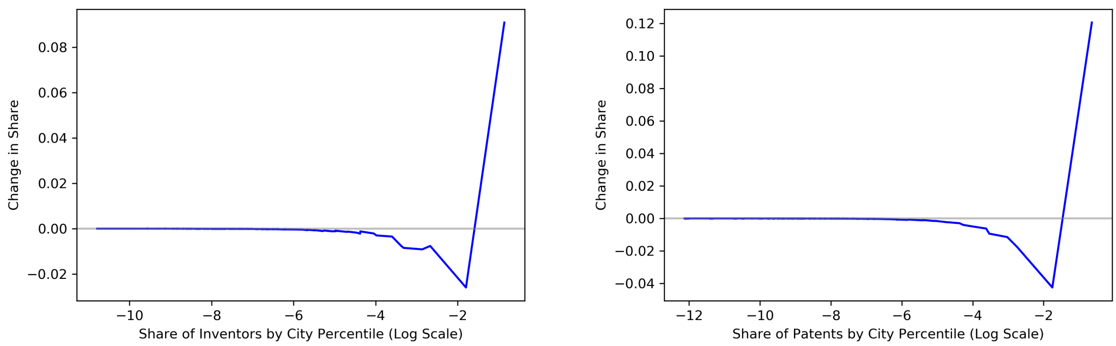
Note: The change in city population is defined as the difference between the average population share of the city in a given decade minus the average population share of the same city between 2000 and 2006.

Figure A.5: Results from Welfare Maximization



(a) Welfare as a function of ξ and ω. (b) Optimal R&D tax credits/subsidies.

Figure A.6: Changes in the spatial distribution of inventors and innovation



(a) Changes in the distribution of inventors (b) Changes in the distribution of innovation

The x-axis shows the share of inventors and patents in each percentile of the city distribution (in log scale), where cities are ordered by their current share of inventors (panel a) or patents filed (panel b). The y-axis plots the expected change in those shares should the economy move to the optimal subsidy scheme. For better visualization, each plot aggregates cities into percentiles.

Table A.1: Calibrated Parameters

Parameter	Value	Description	Matches
ψ	0.5	Elast. innovation wrt R&D	Literature (see footnote 15)
λ	0.132	Innovation step size	Acemoglu et al. (2018)
s_c	$[0.13, 0.30]$	Effective R&D credit rate	Wilson (2009)
ρ	0.02	Discount rate	Literature
g	0.02	Growth rate	Annualized growth rate
r	0.038	Real interest rate	Avg. real interest rate
ε	0.15	Elast. quality in final goods	Profit/sales ratio (BEA)
θ	0.6	Preference parameter	Share of expenditure in non-tradables (BLS; Bems, 2008)
I	1	Population of inventors	Avg. number of inventors residing in CBSA's (1998 - 2016)
L	175	Population of production workers	Avg. employed population residing in CBSA's (1998 - 2016).
L_0	0.85	Population of production workers in CBSA's that do not innovate	Avg. employed population residing in CBSA's w/ no patents filed (1998 - 2016).

Table A.2: Relationship between the R&D credit rate and income taxes.

	Statutory rates		Effective rates		Changes in rates	
	(1)	(2)	(3)	(4)	(5)	(6)
Labor income tax	0.031	0.044	0.085	0.085	-0.071	-0.063
	(0.198)	(0.119)	(0.160)	(0.117)	(0.052)	(0.057)
lag						0.029
						(0.054)
lead						-0.145
						(0.119)
Corporate income tax	0.180	-0.108	-0.081	-0.329	-0.017	0.008
	(0.413)	(0.324)	(0.383)	(0.316)	(0.079)	(0.075)
lag						-0.022
						(0.060)
lead						0.079
						(0.122)
Fixed Effects	State	State, year	State	State, year	-	-
Observations	1530	1530	1530	1530	1479	1377

Standard errors are clustered at the state level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively.

Notes: Columns (1) and (2) regress the statutory R&D tax credit rate on corporate and labor income tax rates. Columns (3) and (4) do the same, but using the effective credit rate. Columns (5) and (6) compute the *within-state* year-on-year changes in statutory R&D credit rates and income tax rates; it then regresses changes in the statutory R&D credit rate on changes in the corporate and labor income tax rates. Column (6) also includes lags and leads of the changes in both tax rates.

Table A.3: Spatial Distribution of Inventors and Patents Over Time

PANEL A: COMPARISON IN LEVELS

Bin	Share of Inventors						Share of Patents					
	1970's		1980's		1990's		1970's		1980's		1990's	
	(model)	(data)	(model)	(data)	(model)	(data)	(model)	(data)	(model)	(data)	(model)	(data)
1	0.002	0.004	0.002	0.004	0.002	0.004	0.001	0.003	0.001	0.002	0.001	0.002
2	0.006	0.01	0.006	0.01	0.006	0.009	0.002	0.004	0.002	0.004	0.002	0.003
3	0.014	0.019	0.014	0.021	0.013	0.018	0.006	0.012	0.006	0.011	0.006	0.009
4	0.041	0.051	0.042	0.054	0.041	0.05	0.02	0.027	0.02	0.031	0.019	0.026
5	0.936	0.916	0.936	0.91	0.938	0.92	0.971	0.954	0.971	0.952	0.972	0.96
Corr.	0.86		0.92		0.96		0.76		0.84		0.95	

PANEL B: REGRESSING DIFFERENCES IN THE DATA ON DIFFERENCES IN THE MODEL

	Δ Share of Inventors			Δ Share of Patents		
	1970's	1980's	1990's	1970's	1980's	1990's
Coefficient	2.27	1.78	2.15	3.27	2.77	3.16
Std. Error	0.09	0.06	0.08	0.08	0.06	0.08
R-squared	0.39	0.45	0.44	0.64	0.68	0.64
Observations	860	860	860	860	860	860

B Proofs and Derivations

Definition 1. *Given initial product quality levels* $\{q_j(0) > 0\}_{j\in\mathcal{J}}$ *and a set of values for the R&D subsidy, effective amenity level and expected productivity of innovation in each city* $\{s_c, \alpha_c, \bar{\chi}_c\}_{c=0}^{C}$ *such that* $\chi_c(t) = \bar{\chi}_c e^{z_c(t)}$, $\bar{\chi}_0 = 0$ *and* $dz_c = \psi(\mu - z_c) + \sigma dW_c$, *a* ***Stationary Balanced Growth Path Equilibrium*** *of the model consists of, for all periods* $t \geq 0$*: (a) an allocation of goods* $\left(Y(t), \{n_c(t)\}_{c=0}^{C}, \{k_j(t)\}_{j\in\mathcal{J}}\right)$*, (b) a spatial distribution of inventors, production workers and firms* $\{I_c(t), L_c(t), N_c(t)\}_{c=0}^{C}$ *and (c) prices* $\{w_c^i(t), w_c^\ell(t), p_{n,c}(t)\}_{c=0}^{C}$ *such that*

(i) $z_c(t)$ *is normally distributed with mean* μ *and variance* σ^2 *for all* t *and all* c*.*

(ii) The production of the final good $Y(t)$*, the average quality of intermediate goods in the economy* $Q(t) = \int_{j\in\mathcal{J}} q_j(t)dj$*, congestion-adjusted wages (*w^i *and* w^ℓ*), and the utility of consumers all grow at a constant rate.*

(iii) All workers are freely mobile and maximize their utility over the consumption of final and non-tradable goods, as well as over the city where they live.

(iv) Final and non-tradable good producers maximize profits while taking prices as given. Intermediate good producers operate under monopolistic competition in the production of each product line.

(v) Incumbent firms take their location as given and choose R&D investments to maximize their discounted stream of profits. There is free entry to all cities and a large mass of potential entrants.

(vi) All labor and goods markets clear, and the amount of public good produced, $\bar{G}$*, balances the government's budget constraint.*

Proof of Lemma 1

In equilibrium, agents are free to choose where they will live. Hence, for any city c and at any time t

$$u_c^h(t) \geq \max_{c'} u_{c'}^h(t) d(c, c'),$$

where $h \in \{i, \ell\}$ denotes the worker type. Without loss of generality, we assume that all cities are populated, so the inequality above must hold for all cities in the economy.

To simplify the notation, let $\delta_c = d_1(c) = 1/d_2(c)$ be the cost/benefit of leaving city c, so $d(c, c') = \delta_c/\delta_{c'}$. Next, we pick an "index" city, c_0, and apply the inequality above (we omit the time notation as it causes no confusion):

$$u_c^h \geq u_{c_0}^h \frac{\delta_c}{\delta_{c_0}};$$

but because city c_0 is also a city, it must be the case that

$$u_{c_0}^h \geq u_c^h \frac{\delta_{c_0}}{\delta_c}.$$

Combining these two equations, we immediately get

$$u_c^h/\delta_c = u_{c0}^h/\delta_{c0} \equiv u^h,$$

where u^h is defined to be the normalized utility of of a type h worker living in city 0. And since city c was arbitrarily picked, the equation above is true for all cities in this economy.

Proof of Lemma 2

City 0. To derive the relationship between population and wages, it is useful to separate city 0 from cities $1, \ldots, C$. The non-tradable good producer in city 0 solves the problem

$$\max_{\ell_{n,0}, m} p_{n,0} n - w_0^\ell \ell_{n,0} - p_{m,0} m \quad \text{s.t. } n = \ell_{n,0}^\beta m^{1-\beta},$$

where w_0^ℓ is determined in equation (2) using the final good producer's problem. The first-order conditions are

$$[\ell_{n,0}]: \quad \beta p_{n,0} \left(\frac{m}{\ell_{n,0}}\right)^{1-\beta} = w_0^\ell$$
$$[m]: \quad (1-\beta) p_{n,0} \left(\frac{\ell_{n,0}}{m}\right)^{\beta} = p_{m,0}.$$

There are three local market clearing conditions (in the sense that they hold inside city 0). The land market clearing condition is

$$m = m_0$$

since land is a fixed factor. The labor market clearing condition is

$$L_0 = \ell_{y,0} + \ell_{n,0},$$

where L_0 is defined as the total population of production workers in city 0. And finally the non-tradable good market clearing condition is

$$\theta \left[\ell_{y,0} \frac{w_0^\ell}{p_{n,0}} + \ell_{n,0} \frac{w_0^\ell}{p_{n,0}}\right] = n$$

where the demand for non-tradable good from each worker is $\theta w_0^\ell / p_{n,0}$, given the familiar Cobb-Douglas utility function of workers.

Using the F.O.C. $[\ell_{n,0}]$ from the non-tradable good producer's problem, the supply of the non-tradable good is

$$n = \frac{w_0^\ell}{p_{n,0}} \frac{\ell_{n,0}}{\beta}.$$

Plugging this into the non-tradable good market clearing condition,

$$\theta\beta[\ell_{y,0} + \ell_{n,0}] = \ell_{n,0} \quad \Longrightarrow \quad \ell_{n,0} = \theta\beta L_0 \quad \text{and} \quad \ell_{y,0} = (1-\theta\beta) L_0.$$

We can also compute the land rent in city 0 by using the F.O.C. $[m]$ and the land market clearing condition:

$$p_{m,0} = (1-\beta)p_{n,0}\left(\frac{\ell_{n,0}}{m_0}\right)^{\beta}.$$

Plug in $p_{n,0}$ from the F.O.C. $[\ell_{n,0}]$ to find

$$p_{m,0}m_0 = (1-\beta)\theta w_0^{\ell}L_0.$$

Now turn to the equilibrium mobility condition $u_0^{\ell}/\delta_0 = u^{\ell}$. A production worker's utility is

$$u^{\ell} = \frac{1}{\delta_0}\left[a_0\left(\theta\frac{w_0^{\ell}}{p_{n,c}}\right)\right]^{\theta}[(1-\theta)w_0^{\ell}]^{1-\theta}.$$

Once again, we can plug in the F.O.C. $[\ell_{n,0}]$ and the labor market clearing condition above to find,

$$u^{\ell} = \left[\alpha_0\theta\beta\left(\frac{m_0}{\theta\beta L_0}\right)^{1-\beta}\right]^{\theta}[(1-\theta)w_0^{\ell}]^{1-\theta},$$

where $\alpha_0 = a_0/\delta_0^{\frac{1}{\theta}}$ is the effective amenity level in city 0. Define $\tilde{L}_0 = L_0/m_0$ as the population per unit of land in city 0. Rearranging the expression above,

$$w_0^{\ell} = \frac{1}{1-\theta}\left[\frac{u^{\ell}}{(\theta\beta)^{\theta\beta}}\right]^{\frac{1}{1-\theta}}\left(\frac{\tilde{L}_0^{1-\beta}}{\alpha_0}\right)^{\frac{\theta}{1-\theta}},$$

as desired.

Cities $1,\ldots,C$**.** The process for cities 1 through C is very similar, with the exception that these cities also have a population of inventors. In each city, the non-tradable good producer solves

$$\max_{\ell_{n,c},m_c} p_{n,c}n - w_c^{\ell}\ell_{n,c} - p_{m,c}m_c \quad \text{s.t. } n = \ell_{n,c}^{\beta}m_c^{1-\beta}.$$

The first-order conditions are

$$[\ell_{n,c}]: \quad \beta p_{n,c}\left(\frac{m_c}{\ell_{n,c}}\right)^{1-\beta} = w_c^{\ell}$$

$$[m_c]: \quad (1-\beta)p_{n,c}\left(\frac{\ell_{n,c}}{m_c}\right)^{\beta} = p_{m,c}.$$

Again, there are three local market clearing conditions that must hold in equilibrium. For all $c \in \{1,\ldots,C\}$, the land market clearing condition is

$$m_c = \bar{m}_c;$$

the labor market clearing condition is

$$\ell_{n,c} = L_c;$$

and the goods market clearing condition is

$$\theta\left[L_c\frac{w_c^\ell}{p_{n,c}}+I_c\frac{w_c^i}{p_{n,c}}\right]=n$$

where L_c and I_c are, respectively, the population of production workers and the population of inventors in city c.

Using the F.O.C. $[\ell_{n,c}]$ and the two latter market clearing conditions, we get

$$\frac{L_c}{\bar{m}_c}=\left(\beta\frac{p_{n,c}}{w_c^\ell}\right)^{\frac{1}{1-\beta}}$$

and

$$\theta\left[L_c\frac{w_c^\ell}{p_{n,c}}+I_c\frac{w_c^i}{p_{n,c}}\right]=\frac{w_c^\ell}{p_{n,c}}\frac{L_c}{\beta}.$$

This second equation simplifies to

$$L_c w_c^\ell=\frac{\theta\beta}{1-\theta\beta}I_c w_c^i. \tag{B.1}$$

The utility level for production workers is therefore

$$\begin{aligned}u^\ell &= \frac{1}{\delta_c}\left[a_c\theta\frac{w_c^\ell}{p_{n,c}}\right]^\theta\left[(1-\theta)w_c^\ell\right]^{1-\theta}\\ &= \left[\alpha_c\theta\beta\left(\frac{\bar{m}_c}{L_c}\right)^{1-\beta}\right]^\theta\left[(1-\theta)w_c^\ell\right]^{1-\theta}.\end{aligned}$$

Rearranging this expression and using the "tilde" to denote variables expressed by units of land ($\tilde{L}_c=L_c/\bar{m}_c$), production worker's wages are

$$w_c^\ell=w^\ell\left(\frac{\tilde{L}_c^{1-\beta}}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\quad\text{where}\quad w^\ell=\frac{1}{1-\theta}\left[\frac{u^\ell}{(\theta\beta)^\theta}\right]^{\frac{1}{1-\theta}}. \tag{B.2}$$

To find w_c^i, and rewrite equation (B.1) as (recall that "tildes" indicate variables per unit of land, $\tilde{I}_c=I_c/\bar{m}_c$)

$$\tilde{L}_c\frac{w_c^\ell}{p_{n,c}}=\frac{\theta\beta}{1-\theta\beta}\tilde{I}_c\frac{w_c^i}{p_{n,c}}.$$

Using $\frac{p_{n,c}}{w_c^\ell}=\frac{\tilde{L}_c^{1-\beta}}{\beta}$, we get

$$\tilde{L}_c=\left[\frac{\theta}{1-\theta\beta}\tilde{I}_c\frac{w_c^i}{p_{n,c}}\right]^{\frac{1}{\beta}}.$$

Now plug this and (B.2) into (B.1) to find

$$\frac{w_c^i}{p_{n,c}}=\frac{1-\theta\beta}{\theta}\left(\frac{1}{w^\ell}\frac{\theta\beta}{1-\theta\beta}w_c^i\right)^{\frac{\beta(1-\theta)}{1-\theta\beta}}\alpha_c^{\frac{\theta\beta}{1-\theta\beta}}\tilde{I}_c^{-\frac{1-\beta}{1-\theta\beta}}.$$

The utility of inventors is thus

$$\begin{aligned}
u^i &= \frac{1}{\delta_c}\left[a_c\theta\frac{w_c^i}{p_{n,c}}\right]^\theta \left[(1-\theta)w_c^i\right]^{1-\theta} \\
&= \left[\alpha_c(1-\theta\beta)\left(\frac{1}{w^\ell}\frac{\theta\beta}{1-\theta\beta}w_c^i\right)^{\frac{\beta(1-\theta)}{1-\theta\beta}}\alpha_c^{\frac{\theta\beta}{1-\theta\beta}}\tilde{I}_c^{-\frac{1-\beta}{1-\theta\beta}}\right]^\theta \left[(1-\theta)w_c^i\right]^{1-\theta}.
\end{aligned}$$

Rearranging,

$$w_c^i = w^i\left(\frac{\tilde{I}_c^{1-\beta}}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} \quad \text{where} \quad w^i = \left\{\frac{u^i}{(1-\theta)^{1-\theta}\left[(1-\theta\beta)\left(\frac{1}{w^\ell}\frac{\theta\beta}{1-\theta\beta}\right)^{\frac{\beta(1-\theta)}{1-\theta\beta}}\right]^\theta}\right\}^{\frac{1-\theta\beta}{1-\theta}}. \tag{B.3}$$

Finally, going back to (B.1) and plugging in (B.2) and (B.3), we get

$$L_c = \left(\frac{\theta\beta}{1-\theta\beta}\frac{w^i}{w^\ell}\right)^{\frac{1-\theta}{1-\theta\beta}} I_c$$

for cities $c \in \{1,\ldots,C\}$. Summing over cities where there is innovation and using that $I = \sum_{c=1}^{C} I_c$ and $L = L_0 + \sum_{c=1}^{C} L_c$, it follows that

$$w^\ell = \left(\frac{I}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}} \frac{\theta\beta}{1-\theta\beta}w^i. \tag{B.4}$$

Finally, plug (B.4) into (B.3) to find

$$w^i = \frac{1}{1-\theta}\left[\frac{u^i}{[(1-\theta\beta)]^\theta}\right]^{\frac{1}{1-\theta}}\left(\frac{I}{L-L_0}\right)^{\frac{\theta\beta}{1-\theta}}.$$

Also note that plugging (B.2) and (B.3) into (B.1) and summing over $c \in \{1,\ldots,C\}$ implies that the number of inventors and production workers is proportional in those cities:

$$\frac{I_c}{I} = \frac{L_c}{L-L_0}.$$

Finally, we can find land rents in each city by plugging in the land market clearing condition and the F.O.C. $[\ell_{n,c}]$ into the F.O.C. $[m_c]$:

$$p_{m,c}\bar{m}_c = \frac{1-\beta}{\beta}w_c^\ell L_c.$$

It is convenient to write this expression in terms of the population and wage of inventors in each city. Using equation (B.1), we have

$$p_{m,c}\bar{m}_c = \frac{(1-\beta)\theta}{1-\theta\beta} w_c^i I_c.$$

□

Proof of Lemma 3

As described in the main text, the firm's HJB equation is

$$rV_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = \max_{x_{f,c}} \left\{ \begin{array}{l} \sum\limits_{q_j \in \mathbf{q}_f} \bar{\pi} L_0 q_j + x_{f,c} \mathbb{E}_j [V_c(\mathbf{q}_f \cup_+ \{(1+\lambda)q_j\}, \tilde{I}_c, Z_c, A) - V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)] \\ \\ -D \sum\limits_{q_j \in \mathbf{q}_f} [V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) - V_c(\mathbf{q}_f \backslash_- \{q_j\}, \tilde{I}_c, Z_c, A)] \\ \\ -(1-s_c) w_c^i (i_{f,c} + \kappa) + \dfrac{\mathbb{E}[dV_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)]}{dt} \end{array} \right\}$$

$$\text{s.t. } x_{f,c} = \bar{\chi}_c Z_c (\tilde{I}_c^{\eta} i_{f,c})^{\psi}$$

where we have defined $Z_c = e^{z_c}$ as the local productivity shock. Since z_c is an Ornstein-Uhlenbeck process with law of motion $dz_c = \phi(\mu - z_c)dt + \sigma dW_c(t)$, it follows that

$$dZ_c = \phi \left(\frac{\sigma^2}{4\phi} - \ln(Z_c) \right) Z_c dt + \sigma Z_c dW_c(t)$$

by application of Itô's lemma and using $\mu = -\frac{\sigma^2}{4\phi}$.

To prove lemma 3, we only need to determine $dV_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)$. This can be done by applying Itô's lemma to the the firm's value function V_c, while taking into account that one of the state variables – the population of inventors per land in the city $\tilde{I}_c$ – is a function of the shock Z_c. For each city $c \geq 1$, define a function $h_c : \mathbb{R}_+ \times \mathbb{R}_+^2 \times [0,1] \times [0,L] \to [0, I/\bar{m}_c]$ such that $\tilde{I}_c = h_c(Z_c; A)$ (recall that $A = (Q, w^i, D, L_0) \in \mathbb{R}_+^2 \times [0,1] \times [0,L]$). Itô's lemma implies that

$$dh_c = \frac{\partial h_c}{\partial A}\frac{\partial A}{\partial t} dt + \left[\phi \left(\frac{\sigma^2}{4\phi} - \ln(Z_c) \right) Z_c \frac{\partial h_c}{\partial Z_c} + \frac{(\sigma Z_c)^2}{2} \frac{\partial^2 h_c}{\partial Z_c^2} \right] dt + \sigma Z_c \frac{\partial h_c}{\partial Z_c} dW_c(t),$$

where the arguments of the function are suppressed for convenience in the notation. Note that the first term in the equation above is the regular derivative of h_c w.r.t. the vector representing the aggregate state of the economy, while the remaining two terms involve differentiating w.r.t. the stochastic process Z_c.

Given the process for h_c, we use Itô's lemma once again to differentiate $V_c(\mathbf{q}_f, h_c(Z_c; A), Z_c, A)$

with respect to time:

$$\begin{aligned} dV_c &= \frac{\partial V_c}{\partial A}\frac{\partial A}{\partial t}dt + \frac{\partial V_c}{\partial Z_c}dZ_c + \frac{\partial V_c}{\partial h_c}dh_c + \frac{1}{2}\left[\frac{\partial^2 V_c}{\partial Z_c^2}(dZ_c)^2 + \frac{\partial^2 V_c}{\partial h_c^2}(dh_c)^2 + 2\frac{\partial^2 V_c}{\partial Z_c \partial h_c}dZ_c dh_c\right] \\ &= \frac{\partial V_c}{\partial A}\frac{\partial A}{\partial t}dt + \frac{\partial V_c}{\partial Z_c}\left[\phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right)Z_c dt + \sigma Z_c dW_c(t)\right] \\ &\quad + \frac{\partial V_c}{\partial h_c}\left\{\left[\phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right)Z_c\frac{\partial h_c}{\partial Z_c} + \frac{(\sigma Z_c)^2}{2}\frac{\partial^2 h_c}{\partial Z_c^2}\right]dt + \sigma Z_c \frac{\partial h_c}{\partial Z_c}dW_c(t) + \frac{\partial h_c}{\partial A}\frac{\partial A}{\partial t}dt\right\} \\ &\quad + \frac{1}{2}\left[\frac{\partial^2 V_c}{\partial Z_c^2}(\sigma Z_c)^2 + \frac{\partial^2 V_c}{\partial h_c^2}\left(\sigma Z_c \frac{\partial h_c}{\partial Z_c}\right)^2 + 2\frac{\partial^2 V_c}{\partial Z_c \partial h_c}(\sigma Z_c)^2 \frac{\partial h_c}{\partial Z_c}\right]dt. \end{aligned}$$

Recall that $\mathbb{E}[dW] = 0$ and note that $\mathbb{E}[\frac{\partial h_c}{\partial A}\frac{\partial A}{\partial t}] = 0$ because the total number of inventors in the economy is fixed. Thus, taking the expectation and replacing $h_c(Z_c; A) = \tilde{I}_c$,

$$\begin{aligned} \frac{\mathbb{E}[dV_c]}{dt} &= \frac{\partial V_c}{\partial A}\frac{\partial A}{\partial t} + \phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right)Z_c\left[\frac{\partial V_c}{\partial Z_c} + \frac{\partial V_c}{\partial \tilde{I}_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right] \\ &\quad + \frac{(\sigma Z_c)^2}{2}\left[\frac{\partial^2 V_c}{\partial Z_c^2} + \frac{\partial^2 V_c}{\partial \tilde{I}_c^2}\left(\frac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + \frac{\partial V_c}{\partial \tilde{I}_c}\frac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + 2\frac{\partial^2 V_c}{\partial Z_c \partial \tilde{I}_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right]. \end{aligned}$$

Finally, define $R_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = \frac{\mathbb{E}[dV_c]}{dt} - \frac{\partial V_c}{\partial A}\frac{\partial A}{\partial t}$ to be the risk that firms face due to the productivity shock Z_c. This concludes the proof.

A note on corporate income taxes The HJB equation above does not include any corporate income taxes, even though these taxes change between locations in the data and are one of the sources of revenue for the government (see section 2). The reason for doing this is that taxing a firm's profits will not change any of its decisions in this model, as long as "profits" include the expenditure on R&D.

To see why, note that the firm's HJB equation including corporate taxes τ_c^π is

$$rv_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = \max_{x_{f,c}} \left\{ \begin{array}{l} (1-\tau_c^\pi)\left[\sum_{q_j \in \mathbf{q}_f} \bar{\pi} L_0 q_j - (1-s_c) w_c^i (i_{f,c} + \kappa)\right] \\ \\ + x_{f,c}\mathbb{E}_j[v_c(\mathbf{q}_f \cup_+ \{(1+\lambda)q_j\}, \tilde{I}_c, Z_c, A) - v_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)] \\ \\ -D \sum_{q_j \in \mathbf{q}_f} [v_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) - v_c(\mathbf{q}_f \backslash_- \{q_j\}, \tilde{I}_c, Z_c, A)] + \frac{\mathbb{E}[dv_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)]}{dt} \end{array} \right\}$$

$$\text{s.t. } x_{f,c} = \bar{\chi}_c Z_c (\tilde{I}_c^\eta i_{f,c})^\psi.$$

For any function v_c that satisfies this equation, we can define $V_c = v_c/(1-\tau_c^\pi)$, where V_c is the solution to the HJB equation shown in the beginning of this proof. The firm's decision of $x_{f,c}$ will thus only depend on the value V_c, which means that firms will choose the same amount of investment in R&D regardless of the corporate income tax rate. In addition, the location decisions of firms will also be independent of this tax, since free entry drives the value of entrants to zero (see the discussion preceding proposition 2).[23]

□

Proof of Proposition 1

We start the proof by solving for the incumbent's value function. To do that, we use a guess ad verify argument. The guess is

$$V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = F \sum_{q_j \in \mathbf{q}_f} q_j + E_c Q,$$

where F and E_c are both functions of the state $(\tilde{I}_c, Z_c, A)$ and $\frac{dF}{dt} = 0$. Substituting this guess along with the constraint into the HJB equation, we find

$$rF \sum_{q_j \in \mathbf{q}_f} q_j + rE_c Q - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t} Q - E_c \dot{Q} =$$

$$\max_{x_{f,c}} \left\{ \begin{array}{l} \bar{\pi} L_0 \sum_{q_j \in \mathbf{q}_f} q_j + x_{f,c} F(1+\lambda)Q - D \sum_{q_j \in \mathbf{q}_f} F q_j \\ \\ -(1-s_c) w_c^i \left(\dfrac{x_{f,c}^{\frac{1}{\psi}}}{(\bar{\chi}_c Z_c)^{\frac{1}{\psi}} \tilde{I}_c^{\eta}} + \kappa \right) \\ \\ +\phi \left(\dfrac{\sigma^2}{4\phi} - \ln(Z_c) \right) Z_c \left[\dfrac{\partial E_c}{\partial Z_c} + \dfrac{\partial E_c}{\partial \tilde{I}_c}\dfrac{\partial \tilde{I}_c}{\partial Z_c} \right] Q + \dfrac{(\sigma Z_c)^2}{2} \left[\dfrac{\partial^2 E_c}{\partial Z_c^2} \right] Q \\ \\ + \dfrac{(\sigma Z_c)^2}{2} \left[\dfrac{\partial E_c}{\partial \tilde{I}_c} \dfrac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \dfrac{\partial^2 E_c}{\partial \tilde{I}_c^2} \left(\dfrac{\partial \tilde{I}_c}{\partial Z_c} \right)^2 + 2 \dfrac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c} \dfrac{\partial \tilde{I}_c}{\partial Z_c} \right] Q \end{array} \right\}$$

The first-order condition is

$$[x_{f,c}]: \quad F(1+\lambda)Q - \frac{1}{\psi} \frac{(1-s_c) w_c^i x_{f,c}^{\frac{1-\psi}{\psi}}}{(\bar{\chi}_c Z_c)^{\frac{1}{\psi}} \tilde{I}_c^{\eta}} \leq 0.$$

If the optimal solution is interior,

$$x_{f,c} = \bar{\chi}_c^{\frac{1}{1-\psi}} \left\{ \psi F (1+\lambda) \frac{Q}{w^i} \frac{\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c} \tilde{I}_c^{\eta - \frac{(1-\beta)\theta}{1-\theta}} \right\}^{\frac{\psi}{1-\psi}} Z_c^{\frac{1}{1-\psi}}$$

where we have used equation (3) to substitute for w_c^i. However, firms might prefer not to invest

[23] Note that an incumbent firm's value will still depend on the corporate tax rate, but the entrant firm's value won't (as it equals zero for all cities).

in R&D at all, particularly in periods where Z_c is low. In those cases, the firm can choose a corner solution $x_{f,c} = 0$, which means that $i_{f,c} = 0$ and it does not need to pay the fixed cost $w_c^i \kappa$ as well. It is useful to analyze each case separately.

<u>Interior Solution.</u> we will start by considering an interior solution. Plugging $x_{f,c}$ in the HJB equation, we find

$$rF \sum_{q_j \in \mathbf{q}_f} q_j + rE_c Q - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t} Q - E_c \dot{Q} =$$

$$\left\{ \begin{array}{l} \bar{\pi} L_0 \sum_{q_j \in \mathbf{q}_f} q_j - D \sum_{q_j \in \mathbf{q}_f} F q_j + (1-\psi) x_{f,c} F (1+\lambda) Q - (1-s_c)\kappa w^i \left(\frac{\tilde{I}_c^{1-\beta}}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} \\ +\phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right) Z_c \left[\frac{\partial E_c}{\partial Z_c} + \frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right] Q + \frac{(\sigma Z_c)^2}{2}\left[\frac{\partial^2 E_c}{\partial Z_c^2}\right] Q \\ +\frac{(\sigma Z_c)^2}{2}\left[\frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \frac{\partial^2 E_c}{\partial \tilde{I}_c^2}\left(\frac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + 2\frac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right] Q, \end{array} \right\}$$

Collecting terms with and without $\sum_{q_j \in \mathbf{q}_f} q_j$,

$$rF = \bar{\pi} L_0 - DF, \tag{B.5}$$

and

$$\begin{aligned} E_c\left(r - \frac{\dot{Q}}{Q}\right) - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t} &= (1-\psi) x_{f,c} F(1+\lambda) - (1-s_c)\kappa \frac{w^i}{Q}\left(\frac{\tilde{I}_c^{1-\beta}}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} \\ &\quad +\phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right) Z_c \left[\frac{\partial E_c}{\partial Z_c} + \frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right] + \frac{(\sigma Z_c)^2}{2}\left[\frac{\partial^2 E_c}{\partial Z_c^2}\right] \qquad \text{(B.6)} \\ &\quad +\frac{(\sigma Z_c)^2}{2}\left[\frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \frac{\partial^2 E_c}{\partial \tilde{I}_c^2}\left(\frac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + 2\frac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right]. \end{aligned}$$

Equation (B.5) immediately gives us

$$F(D, L_0) = \frac{\bar{\pi} L_0}{r + D}. \tag{B.5'}$$

To show that $\frac{dF}{dt} = 0$, it suffices to prove that $\dot{L}_0 = 0$ and $\dot{D} = 0$. Going back to the final good producer's problem and plugging in the quantity of each intermediate good produced, k_j, and the number of production workers used in the production of the final good ($\ell_{y,0} = (1-\theta\beta)L_0$), we find that

$$Y = \frac{1-\theta\beta}{1-\varepsilon}\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}} Q L_0 \quad \Longrightarrow \quad \frac{\dot{Y}}{Y} = \frac{\dot{Q}}{Q} + \frac{\dot{L}_0}{L_0}.$$

By definition, Y and Q grow at a constant rate in a SBGP. From the equation above, this implies that L_0 must either be constant ($\dot{L}_0 = 0$) or grow at a constant rate as well. However, given that the population of production workers is fixed at L, it follows that the only possible rate of growth for L_0 is 0. This argument also implies that Y and Q grow at the same rate, so that $\dot{Y}/Y = \dot{Q}/Q = g$. Furthermore, it will be shown in proposition 3 that the rate of growth of the economy in the SBGP is $g = \lambda D$. Since both g and λ are constant, it follows that D must be constant as well. In conclusion, both L_0 and D are fixed over time, and therefore it follows that the same is true for the franchise value $F(D, L_0)$.

Using $\dot{Q}/Q = g$ and plugging in $x_{f,c}$ from the F.O.C. of the incumbent's problem, (B.6) can be rewritten as

$$\begin{aligned}(r-g)E_c - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t} &= (1-\psi)\bar{\chi}_c[F(1+\lambda)]^{\frac{1}{1-\psi}}\left[\psi\frac{Q}{w^i}\frac{\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c}\tilde{I}_c^{\frac{\eta(1-\theta)-(1-\beta)\theta}{1-\theta}}\right]^{\frac{\psi}{1-\psi}} Z_c^{\frac{1}{1-\psi}} \\ &\quad -(1-s_c)\kappa\frac{w^i}{Q}\left(\frac{\tilde{I}_c^{1-\beta}}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} + \phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right) Z_c\left[\frac{\partial E_c}{\partial Z_c} + \frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right] \\ &\quad + \frac{(\sigma Z_c)^2}{2}\left[\frac{\partial^2 E_c}{\partial Z_c^2} + \frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \frac{\partial^2 E_c}{\partial \tilde{I}_c^2}\left(\frac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + 2\frac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right],\end{aligned} \tag{B.6'}$$

which implicitly defines $E_c = E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)$. To understand the requirement that $r > g$, assume for a moment that $Z_c = 1$ for all cities and in all periods. In this case, equation (B.6') becomes an ordinary differential equation and has an explicit solution:

$$\begin{aligned}E_c(\tilde{I}_c(t), w^i(t)/Q(t), D, L_0) &- \lim_{\hat{t}\to\infty} E_c(\tilde{I}_c(\hat{t}), w^i(\hat{t})/Q(\hat{t}), D, L_0)e^{-(r-g)\hat{t}} = \\ &\int_t^{\infty} e^{-(r-g)(s-t)}\left[\mathcal{K}_c(s)\tilde{I}_c(s)^{\frac{\eta(1-\theta)-(1-\beta)\theta}{1-\theta}\frac{\psi}{1-\psi}} - \mathcal{M}_c(s)\tilde{I}_c(s)^{\frac{(1-\beta)\theta}{1-\theta}}\right] ds,\end{aligned}$$

where $\mathcal{K}_c$ and $\mathcal{M}_c$ are the collection of terms multiplying $\tilde{I}_c$ on the first and second lines of (B.6'), respectively. Since the total population of workers is fixed and $w^i(t)/Q(t)$ is stationary (see proposition 3), the requirement that $r > g$ is sufficient for the limit in the equation above to be zero and for the value of the integral to be well defined. Intuitively, if $g > r$, firms will always find it profitable to invest as much as possible in R&D (by, for example, borrowing capital at rate r), as the value of doing so grows at rate g. This intuitive argument applies to the case with local shocks as well.

<u>Corner Solution.</u> Now let us consider the case of a corner solution. The argument for used

to derive the expression for F remains unchanged; however, equation (B.6) now becomes

$$
\begin{aligned}
rE_c - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t} - E_c\frac{\dot{Q}}{Q} = \;& \phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right) Z_c \left[\frac{\partial E_c}{\partial Z_c} + \frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right] + \frac{(\sigma Z_c)^2}{2}\left[\frac{\partial^2 E_c}{\partial Z_c^2}\right] \\
& + \frac{(\sigma Z_c)^2}{2}\left[\frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \frac{\partial^2 E_c}{\partial \tilde{I}_c^2}\left(\frac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + 2\frac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right]
\end{aligned}
$$

since $x_{f,c} = 0$ and the firm does not have to pay the fixed cost $w_c^i \kappa$. Note that a trivial solution for this equation is $E_c = 0$.

Summarizing both cases, we can conclude that

$$
V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = F(D, L_0)\sum_{q_j \in \mathbf{q}_f} q_j + \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)Q\right\},
$$

where F is given by equation (B.5') and E_c is given by (B.6').

Entrant's Problem. The second stage of the entrant's problem can be solved in the same way, and in particular note that the guess $V_c^e(\tilde{I}_c, Z_c, A) = E_c Q$ implies

$$
\begin{aligned}
& rE_cQ - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t}Q - E_c\dot{Q} = \\
& \max_{x_{f,c}} \left\{
\begin{array}{l}
x_{f,c}F(1+\lambda)Q - (1-s_c)w_c^i\left(\dfrac{x_{f,c}^{\frac{1}{\psi}}}{(\bar{\chi}_c Z_c)^{\frac{1}{\psi}}\tilde{I}_c^{\eta}} + \kappa\right) \\
+\phi\left(\dfrac{\sigma^2}{4\phi} - \ln(Z_c)\right) Z_c\left[\dfrac{\partial E_c}{\partial Z_c} + \dfrac{\partial E_c}{\partial \tilde{I}_c}\dfrac{\partial \tilde{I}_c}{\partial Z_c}\right] Q + \dfrac{(\sigma Z_c)^2}{2}\left[\dfrac{\partial^2 E_c}{\partial Z_c^2}\right] Q \\
+\dfrac{(\sigma Z_c)^2}{2}\left[\dfrac{\partial E_c}{\partial \tilde{I}_c}\dfrac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \dfrac{\partial^2 E_c}{\partial \tilde{I}_c^2}\left(\dfrac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + 2\dfrac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c}\dfrac{\partial \tilde{I}_c}{\partial Z_c}\right] Q
\end{array}
\right\}
\end{aligned}
$$

The first-order condition for this problem is exactly the same as in the incumbent's problem, so the arrival rates of innovation per product line of all firms is the same. Following the same steps as above, it is straightforward to show that the entrant's HJB equation will satisfy equation (B.6'), verifying the guess. The same holds true in case of a corner solution, where the firm hires no inventors and again $E_c = 0$. Collecting both cases, we have that $V_c^e(\tilde{I}_c, Z_c, A) = \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)Q\right\}$, with E_c defined exactly as it was for the incumbent firm. □

Proof of Proposition 2

There are several claims made in proposition 2, so we proceed in order. We start by proving that the ratio w^i/Q does not depend on any of the local shocks Z_c, which readily delivers the population of inventors in each city (equation 6). Then, using those results, we derive the expressions for the number of inventors hired by each firm, the optimal arrival rates of innovation and the number of active firms in each city. Finally, we demonstrate how to find the population of production workers in city 0.

Wages and the Population of Inventors. Imposing the free entry condition ($E_c = 0\ \forall c, t$) into equation (B.6') gives us

$$0 = (1-\psi)\left[\bar{\chi}_c F(1+\lambda)\right]^{\frac{1}{1-\psi}} \left[\psi \frac{Q}{w^i} \frac{\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c} \tilde{I}_c^{\frac{\eta(1-\theta)-(1-\beta)\theta}{1-\theta}}\right]^{\frac{\psi}{1-\psi}} Z_c^{\frac{1}{1-\psi}} - (1-s_c)\kappa \frac{w^i}{Q}\left(\frac{\tilde{I}_c^{1-\beta}}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}.$$

Define $\Theta = (1-\beta)\theta - \psi\eta(1-\theta)$ to be the net elasticity of congestion. Solving for $\tilde{I}_c$ in the equation above yields

$$\tilde{I}_c = \left\{\psi^{\psi}\left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{\bar{\chi}_c(1+\lambda)\bar{\pi}L_0}{r+D} \frac{Q}{w^i} \frac{\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c} Z_c\right\}^{\frac{1-\theta}{\Theta}}. \tag{B.7}$$

Next, we impose that labor markets clear in all periods. The population of inventors in the economy is fixed and equal to I; similarly, the population of production workers equals L. Since workers are freely mobile and supply one unit of labor inelastically, the labor market clearing conditions are (since there are no inventors in city 0)

$$I = \sum_{c=1}^{C} I_c \quad \text{and} \quad L = L_0 + \sum_{c=1}^{C} L_c.$$

Recall that $\tilde{I}_c = I_c/\bar{m}_c$ and that the land mass in cities $c \in \{1, \dots, C\}$ has been normalized to $\bar{m}_c = 1/C$. Multiplying both sides of equation (B.7) by $\bar{m}_c$ and summing over c yields (after rearranging)

$$\frac{w^i}{Q} = \frac{1}{I^{\frac{\Theta}{1-\theta}}}\psi^{\psi}\left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{(1+\lambda)\bar{\pi}L_0}{r+D}\left\{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c}Z_c\right)^{\frac{1-\theta}{\Theta}}\right\}^{\frac{\Theta}{1-\theta}}.$$

Using the expression above, my goal is now to demonstrate that $\frac{w^i}{Q}$ is not a function of the local shocks z_c. For each t, define $\xi_c(t) = \left[\frac{\bar{\chi}_c\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c}Z_c(t)\right]^{\frac{1-\theta}{\Theta}}$. Recall that z_c follows an Ornstein-Uhlenbeck process with stationary distribution $z_c \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2\phi}\right)$. Since $Z_c = e^{z_c}$ and $\mu = -\sigma^2/4\phi$, it follows that Z_c is log-normally distributed with mean 1 and variance $\exp\left(\frac{\sigma^2}{2\phi}\right) - 1$. As a result, $\{\xi_c(t)\}_{c=1}^{C}$ is a sequence of independent random variables, each with finite first and second moments, for all $t \geq 0$. Applying Kolmogorov's Strong Law of Large Numbers (see Shiryaev, 1996, Ch. 4, §3),

$$\frac{1}{C}\sum_{c=1}^{C}\xi_c - \frac{1}{C}\mathbb{E}\left[\sum_{c=1}^{C}\xi_c\right] \xrightarrow{a.s.} 0.$$

To compute the expectation above, note that $(\bar{\chi}_c, \alpha_c, s_c)$ are fixed and Z_c is independent and identically distributed across all c. Furthermore, for any $K \neq 0$, we have

$$\mathbb{E}\left[Z_c^K\right] = \mathbb{E}\left[e^{Kz_c}\right] = \frac{1}{\sqrt{\pi\sigma^2/\phi}} \int_{-\infty}^{\infty} e^{Kz - \frac{1}{2}(z-\mu)^2/\left(\frac{\sigma^2}{2\phi}\right)} dz = e^{K\left(\mu + K\frac{\sigma^2}{4\phi}\right)}.$$

Plugging in $K = (1-\theta)/\Theta$ and $\mu = -\sigma^2/4\phi$, we get $\mathbb{E}\left(Z_c^{\frac{1-\theta}{\Theta}}\right) = \exp\left(\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta} - 1\right)\frac{\sigma^2}{4\phi}\right)$.

Since $C \to \infty$, it follows that wages in each period are, with probability 1,

$$\frac{w^i}{Q} = \frac{1}{I^{\frac{\Theta}{1-\theta}}} \psi^\psi \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{(1+\lambda)\bar{\pi}L_0}{r+D} \left\{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}} e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}\right\}^{\frac{\Theta}{1-\theta}}. \tag{B.8}$$

In words, the equation above shows that w^i/Q does not depend on the individual realizations of the local shocks. It also implies equation (9) in the proposition's statement. Plugging (B.8) into (B.7), the population of inventors in each city is

$$\tilde{I}_c = I \times \frac{\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}}{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}} \times \frac{Z_c^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}, \tag{B.9}$$

which is equivalent to equation (6).

Firms, Inventors per Firm and the Arrival Rate of Innovation. Once wages and the population of inventors is determined in each city, the remaining variables of the model can be readily computed. To begin, simply plug in equation (B.8) into the expression for the optimal arrival rate of innovation (see the proof of proposition 1) to find

$$x_{f,c} = \left[I \frac{\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}}{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}} \frac{Z_c^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}\right]^{\frac{\Theta}{1-\theta}\frac{\psi}{1-\psi}} \psi^\psi \left(\frac{\kappa}{1-\psi}\right)^{\psi} \bar{\chi}_c Z_c \tilde{I}_c^{\frac{\eta(1-\theta)-(1-\beta)\theta}{1-\theta}\frac{\psi}{1-\psi}}.$$

Now use (B.9) to get

$$x_{f,c} = \psi^\psi \left(\frac{\kappa}{1-\psi}\right)^{\psi} \bar{\chi}_c Z_c \tilde{I}_c^{\psi\eta}.$$

The number of inventors hired by each firm is found by substituting the expression for $x_{f,c}$ above into the production function of innovation (1):

$$i_{f,c} = \frac{\psi}{1-\psi}\kappa.$$

Finally, we must also require that all inventors living in the city are employed by some firm. Let $N_c(t)$ be the number of firms located in city c who invest in R&D during period t. Since all firms located in the same city will hire the same number of inventors (in case they decide to invest in R&D), we have that

$$I_c = N_c(i_{f,c} + \kappa).$$

Using the expression for $i_{f,c}$, this expression becomes

$$I_c = \left(\frac{\kappa}{1-\psi}\right) N_c,$$

which equation proves (8) after rearranging.

Population in City 0. Finally, we can now determined the population of city 0. To find L_0, note that equation (B.8) implies

$$w^i = QL_0 \underbrace{\frac{1}{I^{\frac{\Theta}{1-\theta}}} \psi^\psi \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{(1+\lambda)\bar{\pi}}{r+D} \left\{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}} e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}\right\}^{\frac{\Theta}{1-\theta}}}_{\overline{W}^i}.$$

Also, recall that

$$w^\ell = \left(\frac{I}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}} \frac{\theta\beta}{1-\theta\beta} w^i \tag{B.4}$$

and

$$w_0^\ell = w^\ell \left(\frac{\tilde{L}_0^{1-\beta}}{\alpha_0}\right)^{\frac{\theta}{1-\theta}} (\theta\beta)^{\frac{(1-\beta)\theta}{1-\theta}}, \tag{5}$$

where $w_0^\ell = \dfrac{\varepsilon}{1-\varepsilon}\left(\dfrac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}} Q$ from equation (2). Combining all of those results,

$$\begin{aligned}
\frac{\varepsilon}{1-\varepsilon}\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}} Q &= w_0^\ell \\
&= (\theta\beta)^{\frac{(1-\beta)\theta}{1-\theta}} w^\ell \left(\frac{\tilde{L}_0^{1-\beta}}{\alpha_0}\right)^{\frac{\theta}{1-\theta}} \\
&= (\theta\beta)^{\frac{(1-\beta)\theta}{1-\theta}} \left(\frac{I}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}} \frac{\theta\beta}{1-\theta\beta} w^i \left(\frac{\tilde{L}_0^{1-\beta}}{\alpha_0}\right)^{\frac{\theta}{1-\theta}} \\
&= \left(\frac{\theta\beta I}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}} \frac{1}{1-\theta\beta} Q\overline{W}^i L_0 \left(\frac{\tilde{L}_0^{1-\beta}}{\alpha_0}\right)^{\frac{\theta}{1-\theta}}.
\end{aligned}$$

Using the normalization $\bar{m}_0 = 1$ (so that $\tilde{L}_0 = L_0$) and rearranging this equation gives

$$\begin{aligned}
\left(\tfrac{L_0}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}} &= \frac{\varepsilon}{1-\varepsilon}\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}} (1-\theta\beta)\alpha_0^{\frac{\theta}{1-\theta}} \left\{(\theta\beta I)^{\frac{1-\theta\beta}{1-\theta}} \overline{W}^i\right\}^{-1} \\
&= \frac{\alpha_0^{\frac{\theta}{1-\theta}}}{\underbrace{(1-\varepsilon)\psi^\psi \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{1+\lambda}{r+D} e^{\frac{\sigma^2}{4\phi}\left(\frac{1-\theta}{\Theta}-1\right)} I^{1+\psi\eta} (\theta\beta)^{\frac{1-\theta\beta}{1-\theta}} \left[\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}\right]^{\frac{\Theta}{1-\theta}}}_{\equiv \Lambda^{\frac{1-\theta\beta}{1-\theta}}}}
\end{aligned} \tag{B.10}$$

and therefore

$$L_0 = \frac{\Lambda}{1+\Lambda} L.$$

Note that this equation implies that if the rate of creative destruction, D, is constant over time, then so is Λ and therefore so is L_0. Since D is constant in the Balanced Growth Path, it follows that the population in city L_0 is fixed and does not react to any of the shocks in other cities as well.

□

Proof of Corollary 1

To prove this corollary, start with the definition of the rate of creative destruction and plug in the expressions for N_c and $x_{f,c}$ found in proposition 2.

$$\begin{aligned} D &= \sum_{c=1}^{C} N_c x_{f,c} \\ &= \left(\frac{1-\psi}{\kappa}\right) \sum_{c=1}^{C} I_c x_{f,c} \\ &= \left(\frac{1-\psi}{\kappa}\right) \frac{1}{C} \sum_{c=1}^{C} \tilde{I}_c x_{f,c} \\ &= \psi^{\psi} \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{1}{C} \sum_{c=1}^{C} \bar{\chi}_c \bar{\tilde{I}}_c^{1+\psi\eta} Z_c \\ &= \psi^{\psi} \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{1}{C} \sum_{c=1}^{C} \bar{\chi}_c \bar{\tilde{I}}_c^{1+\psi\eta} \frac{Z_c^{1+(1+\psi\eta)\frac{1-\theta}{\Theta}}}{e^{(1+\psi\eta)\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}} \end{aligned}$$

where the third equality uses the fact that $\tilde{I}_c/C = I_c$. The last equality is obtained by plugging in the expression for $\tilde{I}_c$ (equation B.9) and defining $\bar{I}_c = I\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}} \left\{\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}\right\}^{-1}$.

Since the number of cities is large ($C \to \infty$), the Law of Large Numbers applies and the average above converges almost surely to its expected value. Therefore, with probability one,

$$D = \psi^{\psi} \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} e^{\frac{(1-\theta)(1+\psi\eta)}{\Theta}\left(\frac{(1-\beta)\theta}{\Theta}+1\right)\frac{\sigma^2}{4\phi}} \frac{1}{C} \sum_{c=1}^{C} \bar{\chi}_c \bar{\tilde{I}}_c^{1+\psi\eta}$$

where the expectation of $Z_c^{1+(1+\psi\eta)\frac{1-\theta}{\Theta}}$ is found using the relationship $\mathbb{E}\left[Z_c^K\right] = e^{K\left(\mu+K\frac{\sigma^2}{4\phi}\right)}$ found in proposition 2, using $K = 1+(1+\psi\eta)\frac{1-\theta}{\Theta} = \frac{1-\theta\beta}{\Theta}$. □

Proof of Proposition 3

(1) Before we begin the proof, recall from corollary 1 that the rate of creative destruction is fixed over time. As a result, $L_0 = L\Lambda/(1+\Lambda)$ is also constant, since Λ does not change between periods (see proposition 2).

The total production of final goods is given by

$$Y = \frac{\ell_{y,0}^{\varepsilon}}{1-\varepsilon} \int_{\mathcal{J}} q_j^{\varepsilon} k_j^{1-\varepsilon} dj$$

where $\ell_{y,0} = (1-\theta\beta)L_0$. By plugging in k_j from equation (**??**) and doing some algebra,

$$Y = \frac{1-\theta\beta}{1-\varepsilon}\left(\frac{1-\varepsilon}{\nu}\right)^{\frac{1-\varepsilon}{\varepsilon}} L_0 Q.$$

Thus,

$$\frac{\dot{Y}}{Y} = \frac{\dot{Q}}{Q}.$$

Since, by definition, $\dot{Y}/Y = g$, it follows that $\dot{Q}/Q = g$.

To prove that Q and w^i grow at the same rate, it suffices to look at equation (B.8) and realize that the RHS of that equation is fixed over time. Thus

$$d\left(\frac{w^i}{Q}\right)/dt = 0 \Longrightarrow \frac{\dot{w}^i}{w^i} = \frac{\dot{Q}}{Q}.$$

Furthermore, from equation (B.4) it is evident that w^ℓ is proportional to w^i, which means that w^ℓ must also grow at rate g. Finally, equations (4) and (3) show that $u^\ell \propto (w^\ell)^{1-\theta}$ and $u^i \propto (w^i)^{1-\theta}$. From the results above, it follows that both u^ℓ and u^i grow at rate $(1-\theta)g$.

(2) To show that $g = \lambda D$, one can analyze the dynamics of Q. For a small interval of time Δ,

$$Q(t+\Delta) = Q(t) + \sum_{c=1}^{C} N_c(t)(\Delta x_{f,c}(t)) \times \lambda Q(t)$$

where $\Delta x_{f,c}(t)$ is the probability that an innovation will be produced by firm f in city c between t and $t+\Delta$ and $\lambda Q(t)$ is the aggregate expected quality gain from that innovation. Rearranging and taking the limit as $\Delta \to 0$, we get

$$\lim_{\Delta\to 0} \frac{Q(t+\Delta)-Q(t)}{\Delta} = \lambda Q(t) \sum_{c=1}^{C} N_c(t) x_{f,c}(t)$$

$$\frac{\dot{Q}}{Q} = \lambda \sum_{c=1}^{C} N_c x_{f,c}.$$

Since $\dot{Q}/Q = g$ and $\sum_{c=1}^{C} N_c x_{f,c} = D$, the result follows. Note that because D and λ are constant, the rate of growth g must also be constant over time.

(3) Once again, this claim can be proved by analyzing the dynamics of Q_c. For a small interval of time Δ,

$$Q_c(t+\Delta) = Q_c(t) + N_c(t)(\Delta x_{f,c}(t)) \times (1+\lambda)Q - (\Delta D)Q_c(t).$$

Taking the limit as $\Delta \to 0$,

$$\dot{Q}_c = (1+\lambda) Q N_c x_{f,c} - D Q_c.$$

Rearranging this expression and multiplying both sides by e^{Dt}, we have that

$$\frac{d}{dt} e^{Dt} Q_c = (1+\lambda) e^{Dt} Q N_c x_{f,c}.$$

Integrating both sides in $t \in [0, T]$,

$$e^{DT} Q_c(T) - Q_c(0) = (1+\lambda) \int_0^T e^{Dt} Q(t) N_c(t) x_{f,c}(t) dt$$

for a given initial condition $Q_c(0)$.

Under the SBGP equilibrium, Q grows at a constant rate so that $Q(t) = Q(0)e^{gt}$. In addition, $N_c(t)x_{f,c}(t)$ has a stationary distribution so its mean is constant over time. Therefore, taking the expectation on both sides results in

$$e^{DT} \mathbb{E}[Q_c(T)] - Q_c(0) = (1+\lambda) Q(0) \mathbb{E}\left[N_c x_{f,c}\right] \frac{1}{D+g} \left[e^{(D+g)T} - 1\right].$$

Now divide both sides by e^{DT} and take the limit as $T \to \infty$ to find[24]

$$\lim_{t\to\infty} \mathbb{E}[Q_c(t)] = \lim_{t\to\infty} \frac{(1+\lambda)}{D+g} \mathbb{E}\left[N_c x_{f,c}\right] Q(t).$$

Finally, using $g = \lambda D$ we have that

$$\lim_{t\to\infty} \mathbb{E}[Q_c(t)] = \lim_{t\to\infty} \frac{\mathbb{E}\left[N_c x_{f,c}\right]}{D} Q(t). \tag{B.11}$$

Moving back to the law of motion for Q_c, take the expectation on both sides to find

$$\mathbb{E}[\dot{Q}_c(t)] = (1+\lambda) Q(t) \mathbb{E}\left[N_c x_{f,c}\right] - D\mathbb{E}[Q_c(t)].$$

Once again take the limit as $t \to \infty$ and use equation (B.11) to find

$$\lim_{t\to\infty} \mathbb{E}[\dot{Q}_c(t)] = (1+\lambda) D \lim_{t\to\infty} \mathbb{E}[Q_c(t)] - D \lim_{t\to\infty} \mathbb{E}[Q_c(t)]$$

or

$$\lim_{t\to\infty} \mathbb{E}[\dot{Q}_c(t)] = g \lim_{t\to\infty} \mathbb{E}[Q_c(t)].$$

To avoid carrying limits in the notation, let T be a time period such that $\forall t > \mathsf{T}$ the relationship above is approximately true. Define the rate of growth of $Q_c(t)$ as $g_c(t) = \dot{Q}_c(t)/Q_c(t)$. Then, for $t > \mathsf{T}$

$$\mathbb{E}[g_c(t) Q_c(t)] = g\mathbb{E}[Q_c(t)]$$

[24]Alternatively, one can impose $Q_c(0) = \frac{(1+\lambda)Q(0)}{D+g} \mathbb{E}\left[N_c x_{f,c}\right]$ to find the same result for all t.

$$\mathbb{E}[g_c(t)]\mathbb{E}[Q_c(t)] + \text{Cov}(g_c(t), Q_c(t)) = g\mathbb{E}[Q_c(t)]$$
$$\mathbb{E}[g_c(t)] = g - \frac{\text{Cov}(g_c(t), Q_c(t))}{\mathbb{E}[Q_c(t)]}.$$

□

C Linear Regressions

C.1 Data

The main dataset we use to derive the empirical results in section 4.2 is the patent data published by the United States Patent and Trademark Office (USPTO). Through the PatentsView platform, the USPTO provides data on the universe of patents registered in the US, including citations made and received by each patent, their industry classification, who are their inventors and who those patents were assigned to (i.e., who owns the patent). We assume the year in which each patent was "produced" is the year when the patent application was filed. As a baseline quality cutoff, we drop all patents who were never granted.

The focus of this paper is on innovation led by firms, so we drop any patent assignee who is not labeled as a corporation from the data. Throughout the text, we use the words assignee and firm interchangeably to refer to the owner of a patent. We attribute a location to each patent by matching it to its assigned owner and assume that the patent was "produced" at the CBSA where the assignee is located. In case a patent has multiple owners, we split that patent into equal shares between each of them. To minimize double counting when an assignee has multiple addresses, we use the patent's inventors locations to select which address is currently used by that firm.

We also adjust the data for the fact that we only observe inventors and firms when they are successful in producing patents. For example, if a firm files a patent 2009 and in 2011 – but not in 2010 – this firm will not be included in the data in 2010. To deal with this selection problem, we "complete" the dataset by adding back the missing inventors and firms in each year. In the case of firms, we do that by determining an entry and exit year for each firm in every CBSA that they appear in the data. Then we create a new observation whenever a year is missing between entry and exit for each CBSA. In this new observation, the firm is assigned zero patents. In the case of inventors, the procedure is similar, but we also match them to the firms they are "working for" (i.e., the assignees of the patents created by that inventor). As was the case for firms, whenever an inventor has multiple addresses, we use the location of the assignee of the most recent patents he or she created to determine which address is currently used. If this process still results in multiple addresses, we assume that the inventor spends an equal fraction of his or her time in each of those places. As a result, the number of inventors hired by a firm and the population of inventors in a city can be non-integer.

The USPTO data also classifies each patent in one of 6 broad categories and 37 subcategories. For each firm and each inventor, we identify the most common subcategory among all patents produced and assign this subcategory as the firm/inventor's industry. In cases when the mode is not unique, we assign that inventor or firm into a separate industry subcategory, "industry 0." This happens for about 10% of all inventors and firms, but this effect is concentrated on inventors/firms who have produced a small number of patents (for example, an inventor who produces a total of two patents, but in different subcategories, will by definition not have a unique subcategory mode). Weighting by patents produced, only approximately 7% of inventors and 2% of firms are assigned to "industry 0."

We also employ the County Business Patterns (CBP) data, published by the Census Bureau. It contains information on the demography and economic activity (employed population, number of establishments, industry classification, etc.) inside every county in the US. Similarly, we use the Zillow Rent Index (ZRI) and the Zillow Home Value Index (ZHVI), both published by Zillow Research, for data on the median rental value and housing price by square foot in each county (respectively). We aggregate this information to the CBSA level using NBER's county to CBSA crosswalk.[25] In the data used for estimation, we focus on the years after 1998, when the CBP switched its industry classification system from SIC to NAICS. However, we occasionally use data that goes further back in time as well – for example in the construction of the instrument in section 4.2.1, which involves lags of employment shares. The two datasets combined contain information on 2,217,577 patents, 1,191,418 inventors and 136,124 firms (assignees), spread out over 860 CBSAs (not all CBSAs include firms/inventors who produced patents) between the years of 1998 and 2016.

C.2 Going From the Model to the Data

The procedure to transform the continuous time flow variables in the model to quantities that are observed in the data is exactly the same in all cases. Here, we detail how to perform this transformation using the production function of innovation, given by equation (1). Normalizing one year to be equal to a time interval with measure one, the expected number of innovations produced by a firm f located in city c during year T, given the sequence of shocks $\{Z_c(t)\}_{T-1}^{T}$, is

$$\int_{T-1}^{T} x_{f,c}(t)dt = \bar{\chi}_c \int_{T-1}^{T} \left[\left(\frac{I_c(t)}{\bar{m}_c} \right)^{\eta} i_{f,c}(t) \right]^{\psi} Z_c(t)dt.$$

Taking logs

$$\log \int_{T-1}^{T} x_{f,c}(t)dt = \psi \log \left(\kappa \frac{\psi}{1-\psi} \right) + \psi\eta \log(\bar{I}_c) + \log \left(\frac{\bar{\chi}_c}{\bar{m}_c^{\psi\eta}} \right) + \log \int_{T-1}^{T} \mathbf{1}\{i_{f,c}(t) > 0 | Z_c(t)\} \left(\frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}} \right)^{\psi\eta} Z_c(t)dt,$$

where we have used that $i_{f,c}(t) = \kappa \frac{\psi}{1-\psi}$ in the periods when firm f decides to invest in R&D (indicated by $\mathbf{1}\{i_{f,c}(t) > 0 | Z_c(t)\}$) and defined $\bar{I}_c$ as in corollary 1. Once again using those results, we find that

$$\log \int_{T-1}^{T} i_{f,c}(t)dt = \log \left(\kappa \frac{\psi}{1-\psi} \right) + \log \int_{T-1}^{T} \mathbf{1}\{i_{f,c}(t) > 0 | Z_c(t)\}dt$$

[25] See `https://www.nber.org/data/cbsa-fips-county-crosswalk.html`. Out of the 929 CBSAs in the US, 917 can be merged with the CBP data. The remaining 12 are in Puerto Rico, which is not included in the CBP.

and

$$\log \int_{T-1}^{T} I_c(t)dt = \log(\bar{I}_c) + \log \int_{T-1}^{T} \frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}} dt.$$

As a result,

$$\begin{aligned} \log \int_{T-1}^{T} x_{f,c}(t)dt = \; & \psi \log \int_{T-1}^{T} i_{f,c}(t)dt + \psi\eta \log \int_{T-1}^{T} I_c(t)dt + \log\left(\frac{\bar{\chi}_c}{\bar{m}_c^{\psi\eta}}\right) \\ & + \log \int_{T-1}^{T} \mathbf{1}\{i_{f,c}(t) > 0 | Z_c(t)\} \left(\frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}\right)^{\psi\eta} Z_c(t)dt \\ & - \psi \log \int_{T-1}^{T} \mathbf{1}\{i_{f,c}(t) > 0 | Z_c(t)\}dt - \psi\eta \log \int_{T-1}^{T} \frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}} dt. \end{aligned}$$

This last equation leads to the regression model

$$\log(x_{f,c,t}) = \psi \log(i_{f,c,t}) + \psi\eta \log(I_{c,t}) + \delta_c + z_{f,c,t},$$

where, slightly abusing notation so that t now represents a year instead of an infinitesimal period, $x_{f,c,t}$ is the number of innovations produced by a firm f located in city c during year t; $i_{f,c,t}$ is the number of inventors hired by firm f during year t (i.e., the average number of inventors hired per period); $I_{c,t}$ is the population of inventors in city c during year t (i.e., the average population of inventors in city c per period); δ_c is a city fixed-effect that captures variation in $\bar{\chi}_c$ and $\bar{m}_c$; and $z_{f,c,t}$ is a shock aggregating the two bottom lines in the previous equation.

C.3 Details on the Estimation of the Elasticity of Agglomeration

C.3.1 Shift-Share Research Designs

This section provides a brief comparison between the "traditional" shift-share research design and the instrument proposed in section 4.2.1. Following the more common shifft-share approach, one can take difference between variables in (12) relative to their values in period t_0 to find

$$y_{c,t} - y_{c,t_0} = \psi\eta[\log(I_{c,t}) - \log(I_{c,t_0})] + (X'_{c,t} - X'_{c,t_0})\Gamma + (\delta_t - \delta_{t_0}) + (z_{c,t} - z_{c,t_0}),$$

where $y_{c,t}$ is the outcome variable in the regression and t_0 is some pre-period, usually many years before year t so that there is no correlation between $z_{c,t}$ and I_{c,t_0}. An instrument for $[\log(I_{c,t}) - \log(I_{c,t_0})]$ could then be defined as

$$\log(\mathcal{I}^{diff}_{c,t,t_0}) = \log \sum_{k=1}^{K} \frac{I_{k,c,t_0}}{I_{c,t_0}}(1 + \gamma_{k,t_0 \to t}).$$

Recall from the definition of $\mathcal{I}_{c,t,l}$ in equation (13) that this is exactly what we would get if we computed the difference $\log(\mathcal{I}_{c,t,l})\big|_{l=t-t_0} - \log(I_{c,t_0})$.

Alternatively, the fixed-effects regression model we run is equivalent to de-meaning the variables in (12):

$$y_{c,t} - \overline{y_c} = \psi\eta[\log(I_{c,t}) - \overline{\log(I_c)}] + (X'_{c,t} - \overline{X'_c})\Gamma + (\delta_t - \overline{\delta}) + (z_{c,t} - \overline{z_c}),$$

where the overlines indicate averages over time. Given the definition of $\mathcal{I}_{c,t,l}$ (eq. 13), the instrument for $[\log(I_{c,t}) - \overline{\log(I_c)}]$ is

$$\log(\mathcal{I}_{c,t,l})\big|_{l=t-t_0} - \overline{\log(I_c)} = \log\sum_{k=1}^{K}\frac{I_{k,c,t_0}}{\hat{I}_c}(1+\gamma_{k,t_0\to t}),$$

where $\log(\hat{I}_c) = \overline{\log(I_c)}$. The difference between the shift-share design and the fixed effects specification therefore resides only on the denominator used to compute the "share." As shown by Adão et al. (2019) and Borusyak et al. (2022), the econometric properties of the estimators using this type of instrument do not depend on the exact definition of these shares. Furthermore, by using the fixed-effects specification, data on the outcome $y_{c,t}$ and the control variables $X_{c,t}$ during period t_0 is not required to estimate the parameters in the regression.

C.3.2 Threats to Identification

The exogeneity condition for the instrument defined in section 4.2.1 is $\mathbb{E}_c\left[\mathcal{I}_{c,t,l} z_{c,t}\right] = 0$. Following the argument made by Borusyak et al. (2022), this condition can be written as

$$\sum_{k=1}^{K} I_{k,t-l}\omega_{k,t}(1+\gamma_{k,t-l\to t}) = 0,$$

where $I_{k,t-l} = \mathbb{E}_c[I_{k,c,t-l}]$ and $\omega_{k,t} = \mathbb{E}_c[I_{k,c,t-l}z_{c,t}]/\mathbb{E}_c[I_{k,c,t-l}]$. This equation provides two different interpretations for the conditions that are required for instrument exogeneity. The first one is that the lagged industry employment levels $I_{k,c,t-l}$ are uncorrelated with the current shock, so $\omega_{k,t} = 0$. As mentioned in the main text, this condition is unlikely to hold for contemporary shares/shocks. However, even is shocks are serially correlated, this condition will hold if lags are long enough. The estimates shown in table 1 cover this case, showing the results when the instrument is lagged for up to 10 years and when industry employment is fixed at its average level between 1990 and 1995.

The second interpretation for instrument endogeneity requires that

$$\sum_{k=1}^{K} I_{k,t-l}\omega_{k,t}(1+\gamma_{k,t-l\to t}) \longrightarrow 0$$

as the number of industries increases. In words, the industry-specific growth rate of employment in countries other than the US, $\gamma_{k,t-l\to t}$, is asymptotically uncorrelated with the industry-specific average of unobserved factors affecting the employment level in locations specializing in each industry, $\omega_{k,t}$. One concern that arises in this case are industries that are simultaneously highly concentrated in specific cities and large enough to drive international trends in employment shares. To avoid this issue, we slightly modify the instrument to exclude any industry whose employment share in any single city exceeds 15% at any point in time.[26] This excludes 6 of the original 38 industries (information storage, drugs, semiconductor devices, motors and engines/parts, apparel and textile, earth working and wells). The estimation results can be seen in table C.4. While 32 industries might be a low number to claim asymptotic lack of correlation, the fact that the estimates below are similar to the ones found before indicates that no city is likely to be driving $\gamma_{k,t-l\to t}$ on its own.

Table C.4: Estimation of the elasticity of agglomeration – excludes spatially concentrated industries.

First Stage	(1)	(2)	(3)	(4)
$\log(\mathcal{I}_{c,t,l})$	0.532***	0.467***	0.389***	0.260***
	(0.038)	(0.040)	(0.039)	(0.041)
F-stat. excluded inst.	196.20	137.93	97.70	40.71

Second Stage	(1)	(2)	(3)	(4)
log(Inventors in City)	0.109***	0.103***	0.092***	0.090
	(0.021)	(0.022)	(0.026)	(0.057)
AKM SE	(0.009)	(0.011)	(0.011)	(0.001)
Method	IV ($l=5$)	IV ($l=7$)	IV ($l=10$)	IV ($l=t-t_{90-95}$)
Observations	11184	11165	11119	11162
Implied η	0.218	0.206	0.184	0.180

Standard errors are clustered at the CBSA level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively. AKM SE indicates alternative standard errors, calculated according to Adão et al. (2019). All specifications control for CBSA and year fixed effects.

C.3.3 Robustness Checks

Including Cities with Zero Patents. Recall from equation (11) that the theoretical model presented in this paper implies the following relationship in the data:

$$\log(\text{patents}_{f,c,t}) - \psi \log(i_{f,c,t}) = \psi\eta \log(I_{c,t}) + X'_{f,c,t}\Gamma + \delta_c + \delta_t + z_{f,c,t}.$$

One issue with this equation is that many firms, especially those in smaller cities, do not produce any patents over the course of one year. If this is the case, running the log-log regression above simply discards those observations, potentially biasing the estimation. To avoid this problem, we can instead run this regression in the form

$$Y_{f,c,t} = \exp\left(\psi\eta \log(I_{c,t}) + X'_{f,c,t}\Gamma + \delta_c + \delta_t + z_{f,c,t}\right),$$

[26]The value of the threshold is arbitrary, but variations around it (e.g., 10, 20 or 25%) generate comparable results.

where

$$Y_{f,c,t} = \begin{cases} \text{patents}_{f,c,t}/i^{\psi}_{f,c,t}, & \text{if } i_{f,c,t} > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Once again, the goal is to estimate the elasticity of agglomeration $\psi\eta$, so we can aggregate the data by taking averages over cities.

$$Y_{c,t} = \exp\left(\psi\eta\log(I_{c,t}) + X'_{c,t}\Gamma + \delta_c + \delta_t + z_{c,t}\right)$$

where $Y_{c,t} = \frac{1}{N_{c,t}}\sum_{f=1}^{N_{c,t}} Y_{f,c,t}$ is the average of outcomes $Y_{f,c,t}$ and $\exp(z_{c,t}) = \frac{1}{N_{c,t}}\sum_{f=1}^{N_{c,t}} \exp(z_{f,c,t})$ aggregates the residuals. For consistency with the other empirical models run so far, $X_{c,t}$ includes the average number of citations that all firms in city c receive in year t and the industry employment shares in city c during year t.

We control for the endogeneity of $\log(I_{c,t})$, by following the method proposed by Wooldridge (2010, Ch.19), which approaches the endogeneity problem from an omitted variables perspective. Specifically, let

$$\log(I_{c,t}) = [\log(\mathcal{I}_{c,t,l}), X'_{c,t}, \delta_c, \delta_t] \cdot \Pi + v_{c,t} \tag{C.1}$$

where Π is vector of reduced-form parameters and $v_{c,t}$ is a residual. Furthermore, we assume that $(z_{c,t}, v_{c,t})$ are independent of $[\log(\mathcal{I}_{c,t,l}), X'_{c,t}, \delta_c, \delta_t]$ and that

$$z_{c,t} = \rho v_{c,t} + e_{c,t}$$

where $e_{c,t}$ is independent of $v_{c,t}$.

Under those conditions, and assuming that $Y_{c,t}$ has a Poisson distribution, it follows that (slightly abusing notation so that the fixed effects absorb constant terms)

$$\log\left(\mathbb{E}\left[Y_{c,t} \middle| I_{c,t}, X_{c,t}, \delta_c, \delta_t, v_{c,t}\right]\right) = \psi\eta\log(I_{c,t}) + X'_{c,t}\Gamma + \delta_c + \delta_t + \rho v_{c,t}. \tag{C.2}$$

Intuitively, the term $\rho v_{c,t}$ controls for the endogeneity of $\log(I_{c,t})$. The residuals $v_{c,t}$ can be obtained from the first-stage regression (C.1) and the model (C.2) is estimated via pseudo-maximum likelihood with the help the `ppmlhdfe` command in Stata, developed by Correia et al. (2019). Once again, each observation is weighted by the number of firms in each city to reflect the fact that the aggregated data is comprised of means over these firms.

The estimated coefficients are in table C.5. Column (1) shows the coefficient estimate without the use of an instrument ($\rho = 0$), columns (2) - (4) show the estimates obtained when using different lags l to compute the instrument and column (5) fixes $I_{k,c,t_{90-95}}$ at its average value between 1990 and 1995. The Poisson model produces larger point estimates than the linear model, potentially because the effect of spillovers at the extensive margin (and on firms who do not continuously innovate) is larger than for other incumbents. However, the standard errors are also quite big, leading values that are not statistically different from zero in columns (4) and (5).[27]

[27]Note that the AKM standard errors are not shown in this case, as the method developed by Adão et al. (2019) only applies to linear regressions.

Table C.5: Estimation of the elasticity of agglomeration – Poisson regression.

First Stage					
	(1)	(2)	(3)	(4)	(5)
$\log(\mathcal{I}_{c,t,l})$		0.536***	0.469***	0.388***	0.248***
		(0.031)	(0.033)	(0.034)	(0.039)
F-stat. excluded inst.		308.28	206.71	129.93	40.02

Second Stage					
log(Inventors in City)	0.197***	0.122***	0.095**	0.029	-0.001
	(0.030)	(0.039)	(0.043)	(0.052)	(0.120)
Method	OLS	IV ($l = 5$)	IV ($l = 7$)	IV ($l = 10$)	IV ($l = t - t_{90-95}$)
Observations	12889	12827	12808	12766	12792
Implied η	0.394	0.244	0.190	0.058	-0.002

Standard errors are clustered at the CBSA level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively. All specifications control for patent quality and city industry composition, as well as CBSA and year fixed effects.

Other Sources of Externality. The population of inventors in a city might not be the only source of agglomeration for firms investing in innovation. For example, an individual firm might benefit from locating near other companies, as this provides opportunities for the firm to learn from its competitors and improve the quality of its own investments. Similarly, the agglomeration spillover, as assumed by a number of papers in the literature, might be related to overall population density instead of the density of inventors.

To test these hypotheses, we include various sources of agglomeration into the regression equation (12): the total population of inventors, the total number of firms who invest in R&D, the total employed population and the total number of (overall) establishments in each city. We focus on simple correlations in this case, so no instrumental variable specification is shown in table C.6 below. Column (1) shows the coefficients from the log-linear model, and column (2) shows the coefficients from the Poisson counting model. As shown in the table, only the number of inventors in the city has a positive (and significant) effect over the average production of patents in each city. This result also holds in the disaggregated model (11), in different geographic levels of aggregation (e.g., counties), and after including non-linear functions of each of the potential sources of agglomeration into the regression model.

C.4 Details on the Estimation of the Elasticity of Congestion

To arrive at the regression model (14) from the equations in the model, we basically follow the argument described in section C.2 of this appendix. Plugging in the value of production workers' wages (eq. 4) into to intermediate good producer's FOC, we find

$$p_{n,c} = \frac{w^{\ell}}{\beta \alpha_c^{\frac{\theta}{1-\theta}}} \left(\frac{L_c}{\bar{m}_c} \right)^{\frac{1-\beta}{1-\theta}}$$

Table C.6: Estimation of the elasticity of agglomeration – multiple sources of agglomeration.

	(1)	(2)
log(Inventors in City)	0.064***	0.203***
	(0.017)	(0.034)
log(Innov. Firms in City)	0.007	-0.002
	(0.014)	(0.029)
log(Employment in City)	-0.078	-0.124
	(0.076)	(0.146)
log(Establishments in City)	0.106*	-0.004
	(0.062)	(0.151)
Model	Log-linear (OLS)	Poisson
Observations	11279	12889

Standard errors are clustered at the CBSA level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively. All specifications control for patent quality and city industry composition, as well as CBSA and year fixed effects.

for cities $c \geq 1$. In city 0, a similar equation holds, but the RHS is multiplied by $(\theta\beta)^{\frac{(1-\beta)\theta}{1-\theta}}$. The equation above reflects the flows of continuous variables. To match them to the data, we integrate them over the period of one year:

$$\int_{T-1}^{T} p_{n,c}(t)dt = \left[\beta\alpha_c^{\frac{\theta}{1-\theta}}\bar{m}_c^{\frac{1-\beta}{1-\theta}}\right]^{-1}\int_{T-1}^{T} w^{\ell}(t)L_c(t)^{\frac{1-\beta}{1-\theta}}dt.$$

In city 0, the population of inventors is constant, so the integral reduces to $\int_{T-1}^{T} w^{\ell}(t)dt$. In cities $c \in \{1,\ldots,C\}$, the population of production workers is proportional to the population of inventors (see lemma 2). As a result,

$$L_c(t) = \frac{L-L_0}{I}I_c(t) = \frac{L-L_0}{I}\bar{I}_c\frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}},$$

where $\bar{I}_c$ is constant over time. Define $\bar{L}_c = \frac{L-L_0}{I}\bar{I}_c$ so that

$$\int_{T-1}^{T} p_{n,c}(t)dt = \left[\beta\alpha_c^{\frac{\theta}{1-\theta}}\bar{m}_c^{\frac{1-\beta}{1-\theta}}\right]^{-1}\bar{L}_c^{\frac{1-\beta}{1-\theta}}\int_{T-1}^{T} w^{\ell}(t)\frac{Z_c(t)^{\frac{1-\beta}{\Theta}}}{e^{\frac{1-\beta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}dt.$$

Take logs to find

$$\log\int_{T-1}^{T} p_{n,c}(t)dt = \left(\frac{1-\beta}{1-\theta}\right)\log(\bar{L}_c) - \log\left(\beta\alpha_c^{\frac{\theta}{1-\theta}}\bar{m}_c^{\frac{1-\beta}{1-\theta}}\right) + \log\int_{T-1}^{T} w^{\ell}(t)\frac{Z_c(t)^{\frac{1-\beta}{\Theta}}}{e^{\frac{1-\beta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}dt.$$

Finally, use the fact that $\int_{T-1}^{T} L_c(t)dt = \bar{L}_c\int_{T-1}^{T}\frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}dt$ to replace $\bar{L}_c$ in the equation above:

$$\log \int_{T-1}^{T} p_{n,c}(t)dt = \left(\frac{1-\beta}{1-\theta}\right) \log \int_{T-1}^{T} L_c(t)dt - \log\left(\beta \alpha_c^{\frac{\theta}{1-\theta}} \bar{m}_c^{\frac{1-\beta}{1-\theta}}\right)$$

$$+ \log \int_{T-1}^{T} w^{\ell}(t) \frac{Z_c(t)^{\frac{1-\beta}{\Theta}}}{e^{\frac{1-\beta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}} dt - \left(\frac{1-\beta}{1-\theta}\right) \log \int_{T-1}^{T} \frac{Z_c(t)^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}} dt,$$

In the expression above, $\int_{T-1}^{T} p_{n,c}(t)dt$ is the average price charged for the non-tradable good in city c during year T, which we approximate by the median rent value per square foot of housing units in city c, $p^h_{c,t}$. Similarly, $\int_{T-1}^{T} L_c(t)dt$ is the average population of production workers living in city c during year T. This is matched to the total employed population minus the number of inventors in each CBSA in the data. The other terms in the equation are captured by a city fixed effect, which accounts for differences in land mass and effective amenities in each city; a year fixed effect, controlling for growth in wages $w^{\ell}(t)$; and a local shock $z^h_{c,t}$ that combines the variation in the integrals containing $Z_c(t)$. This produces the model (14),

$$\log(p^h_{c,t}) = \left(\frac{1-\beta}{1-\theta}\right) \log(L_{c,t}) + \delta_c + \delta_t + z^h_{c,t}.$$

Note that $L_{c,t}$ is correlated with the shock $z^h_{c,t}$, since $L_c(t)$ varies with the shock $z_c(t)$.

C.4.1 Threats to Identification

Following the discussion in sections 4.2.2 and C.3.2, table C.7 shows the estimated elasticity of rental prices with respect to the city's population of production workers. For the estimation of those elasticities, the instrument $\mathcal{I}_{c,t,l}$ is computed after excluding industries whose employment share in any single city exceeds 15% in any point in time. The implied value of β does not show any significant change when compared to table 2.

C.4.2 Alternative Specification

Rental value data is only available in the Zillow Rent Index starting late in 2010. To take advantage of a larger dataset, we run an alternative regression that approximates the price of the non-tradable good in each city by the median price per square foot of housing units. This series is available since 1996 in the Zillow Home Value Index database. The main issue with using housing prices as an approximation for the price of non-tradable goods (whose consumption is modeled as a flow) is that houses are long-term assets, and therefore their prices could be influenced by agents' expectations about the future. Having that in mind, housing prices can still provide useful information on congestion costs in each city.

Table C.8 displays the estimates elasticity of housing prices with respect to the city's population of production workers. The estimated coefficients are quite smaller than in table 2, and imply a value of β closer to 0.8. These numbers reflect a lower elasticity of housing prices (relative to rental prices) to the city's population, most likely because those prices only reflect permanent changes in congestion costs in each city. The value of β found using this alternative

Table C.7: Estimation of the elasticity of congestion – excludes spatially concentrated industries.

First Stage

	(1)	(2)	(3)	(4)
$\log(\mathcal{I}_{c,t,l})$	0.013***	0.017***	0.013***	0.022***
	(0.004)	(0.007)	(0.005)	(0.007)
F-stat. excluded inst.	11.20	5.67	8.31	9.56

Second Stage

	(1)	(2)	(3)	(4)
log(Prod. Workers in City)	1.336***	1.063**	1.384***	1.033***
	(0.413)	(0.459)	(0.479)	(0.368)
AKM SE	(0.019)	(0.021)	(0.032)	(0.006)
Method	IV ($l = 5$)	IV ($l = 7$)	IV ($l = 10$)	IV ($l = t - t_{90-95}$)
Observations	2843	2833	2822	2825
Implied β	0.466	0.575	0.446	0.587

Standard errors are clustered at the CBSA level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively. AKM SE indicates alternative standard errors, calculated according to Adão et al. (2019). All specifications control for CBSA and year fixed effects.

Table C.8: Estimation of the elasticity of congestion – housing prices.

First Stage

	(1)	(2)	(3)	(4)
$\log(\mathcal{I}_{c,t,l})$	0.039***	0.037***	0.035***	0.032***
	(0.006)	(0.006)	(0.006)	(0.007)
F-stat. excluded inst.	42.94	36.00	40.04	22.42

Second Stage

	(1)	(2)	(3)	(4)
log(Prod. Workers in City)	0.335*	0.322*	0.418**	0.426**
	(0.183)	(0.179)	(0.166)	(0.208)
AKM SE	(0.067)	(0.060)	(0.059)	(0.002)
Method	IV ($l = 5$)	IV ($l = 7$)	IV ($l = 10$)	IV ($l = t - t_{90-95}$)
Observations	9900	9891	9874	9876
Implied β	0.866	0.871	0.833	0.830

Standard errors are clustered at the CBSA level and shown in parenthesis. *, **, and *** indicate that the coefficient is statistically different from 0 at the 10%, 5%, and 1% levels, respectively. AKM SE indicates alternative standard errors, calculated according to Adão et al. (2019). All specifications control for CBSA and year fixed effects.

specification is used as an upper bound in sensitivity analyses of the counterfactual results in this paper.

D Matching Moments

D.1 Identifying α_0 and the Scale of χ_c

Identifying the scale of the mean productivity in each city and the effective amenity level in city 0 is straightforward once $\sigma^2/4\phi$ (along with the remaining parameters in the model) is known. For now, assume that this is the case. Define the scale of productivity as χ^s, where $\bar{\chi}_c = \chi^s\hat{\chi}_c$ and $\mathbb{E}[\hat{\chi}_c] = 1$. Then, from corollary 1, we have that

$$D = \chi^s \psi^\psi \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} e^{\frac{(1-\theta)(1+\psi\eta)}{\Theta}\left(\frac{(1-\beta)\theta}{\Theta}+1\right)\frac{\sigma^2}{4\phi}} \frac{1}{C}\sum_{c=1}^{C} \hat{\chi}_c \tilde{\bar{I}}_c^{1+\psi\eta}.$$

Given that $g = \lambda D$, it follows that

$$\chi^s = \frac{g}{\lambda}\left\{\psi^\psi \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} e^{\frac{(1-\theta)(1+\psi\eta)}{\Theta}\left(\frac{(1-\beta)\theta}{\Theta}+1\right)\frac{\sigma^2}{4\phi}} \frac{1}{C}\sum_{c=1}^{C} \hat{\chi}_c \tilde{\bar{I}}_c^{1+\psi\eta}\right\}^{-1}.$$

To identify α_0, recall from the proof of proposition 2 that

$$\left(\frac{L_0}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}} = \frac{\alpha_0^{\frac{\theta}{1-\theta}}}{(1-\varepsilon)\psi^\psi \left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{1+\lambda}{r+D} e^{\frac{\sigma^2}{4\phi}\left(\frac{1-\theta}{\Theta}-1\right)} I^{1+\psi\eta}(\theta\beta)^{\frac{1-\theta\beta}{1-\theta}} \left[\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}\right]^{\frac{\Theta}{1-\theta}}}.$$

The LHS of the equation above can be constructed using the data and values of the parameters already estimated. The same is true for the denominator in the RHS, since $\bar{\chi}_c$ is now fully determined. The equation above therefore identifies α_0, given the values of parameters estimated in the previous steps.

D.2 Law of Motion of the Productivity Shock

Finally, in this section we describe the identification of the parameters in the stochastic process of the productivity shock, Z_c. As mentioned in the main text, only the ratio σ^2/ψ influences the values of the variables of interest in the SBGP equilibrium, so we set $\phi = 1$. The remaining parameter, σ, is then identified by the cross-sectional variance of the population of inventors across cities. Recall that quantities in the model must be integrated over time to match the same frequency as the data. The variance of the population of inventors across cities in year T is therefore

$$\begin{aligned}
\text{Var}(I_{c,T}) &= \frac{1}{C}\sum_{c=1}^{C}\mathbb{E}\left[\left(\int_{T-1}^{T} I_c(t)dt\right)^2\right] - \mathbb{E}\left[\int_{T-1}^{T} I_c(t)dt\right]^2 \\
&= \frac{1}{C}\sum_{c=1}^{C}\mathbb{E}\left[\int_{T-1}^{T} I_c(t)dt \int_{T-1}^{T} I_c(s)ds\right] - \mathbb{E}\left[\int_{T-1}^{T} I_c(t)dt\right]\mathbb{E}\left[\int_{T-1}^{T} I_c(s)ds\right] \\
&= \frac{1}{C}\sum_{c=1}^{C}\int_{T-1}^{T}\int_{T-1}^{T}\left\{\mathbb{E}[I_c(t)I_c(s)] - \mathbb{E}[I_c(t)]\mathbb{E}[I_c(s)]\right\}dtds \\
&= \frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{I}_c}{\exp\left(\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}\right)}\right)^2\int_{T-1}^{T}\int_{T-1}^{T}\text{Cov}\left(Z_c(t)^{\frac{1-\theta}{\Theta}}, Z_c(s)^{\frac{1-\theta}{\Theta}}\right)dtds
\end{aligned}$$

To compute this covariance, we start by noting that the limiting (stationary) distribution of the Ornstein-Uhlenbeck process is such that $\text{Cov}(z_c(t), z_c(s)) = \frac{\sigma^2}{2\phi}e^{-\phi|t-s|}$. Using the properties of the multi-variate log-normal distribution, it follows that $\text{Cov}\left(Z_c(t)^{\frac{1-\theta}{\Theta}}, Z_c(s)^{\frac{1-\theta}{\Theta}}\right) = \exp\left(\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{2\phi}\right)\left[\exp\left(\frac{1-\theta}{\Theta}\frac{\sigma^2}{2\phi}\exp(-\phi|s-t|)\right)-1\right]$. Plugging into the equation above,

$$\text{Var}(I_{c,T}) = \frac{1}{C}\sum_{c=1}^{C}\bar{I}_c^2\left[\int_0^1\int_0^1\exp\left(\frac{1-\theta}{\Theta}\frac{\sigma^2}{2\phi}\exp(-\phi|s-t|)\right)dtds - 1\right],$$

where a simple change in variables switches the region of integration to $[0,1]$.[28] The LHS of the equation above is the average cross-sectional variance of the population of inventors across cities, which can be computed in the data. Given $\phi = 1$ and the values of the other parameters already identified, this equation identifies σ.

E The Government's Problem

The government's problem, as stated in section 5, is

$$\max_{\{s_c\}_{c=1}^{C}} \quad \int_0^\infty e^{-\rho t}\left\{\sum_{c=0}^{C}\left[L_c(t)u^\ell(t) + I_c(t)u^i(t)\right]\right\}G(t)dt$$

$$\text{s.t.} \quad \int_0^\infty e^{-rt}\left[\sum_{c=1}^{C}s_c w_c^i(t)I_c(t) + \gamma(G(t))\right]dt = \int_0^\infty e^{-rt}\left[p_{m,0}(t)\bar{m}_0 + \sum_{c=1}^{C}p_{m,c}(t)\bar{m}_c + \Pi(t)\right]dt$$

[28]The integral in this equation does not have a closed form solution, but it can be simplified to $2\int_{e^{-\phi}}^{1}\int_{1}^{1/y}\exp\left(\frac{1-\theta}{\Theta}\frac{\sigma^2}{2\phi}xy\right)\frac{1}{\phi^2 xy}dxdy$, which makes the numerical integration easier.

The sum in the objective function reduces to $Lu^\ell(t)+Iu^i(t)$ because market clearing must hold in all periods. Using equations (3), (4) and (B.4), this term can be rewritten as

$$(L-L_0)^{\theta\beta}I^{1-\theta\beta}\left(1+\theta\beta\frac{L_0}{L-L_0}\right)\left(\frac{1-\theta}{1-\theta\beta}w^i(t)\right)^{1-\theta}.$$

Using the results in lemma 2, the term $p_{m,0}(t)\bar{m}_0+\sum_{c=1}^{C}p_{m,c}(t)\bar{m}_c$ in the budget constraint is

$$p_{m,0}(t)\bar{m}_0+\sum_{c=1}^{C}p_{m,c}(t)\bar{m}_c=(1-\beta)\theta w_0^\ell(t)L_0+\sum_{c=1}^{C}\frac{(1-\beta)\theta}{1-\theta\beta}w_c^i(t)I_c(t).$$

From equations (5) and (B.4), the first part of the expression above is

$$\begin{aligned}(1-\beta)\theta w_0^\ell(t)L_0 &\overset{(5)}{=} (1-\beta)\beta^{\frac{(1-\beta)\theta}{1-\theta}}\left(\frac{1}{\alpha_0}\right)^{\frac{\theta}{1-\theta}}(\theta L_0)^{\frac{1-\theta\beta}{1-\theta}}w^\ell(t)\\ &\overset{(B.4)}{=} \frac{(1-\beta)\theta}{1-\theta\beta}\left(\frac{1}{\alpha_0}\right)^{\frac{\theta}{1-\theta}}\left(\theta\beta\frac{L_0}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}}I^{\frac{1-\theta\beta}{1-\theta}}w^i(t).\end{aligned}$$

Equation (3) and the definition $I_c=\tilde{I}_c/C$ yield

$$\sum_{c=1}^{C}\frac{(1-\beta)\theta}{1-\theta\beta}w_c^i(t)I_c(t)=\frac{(1-\beta)\theta}{1-\theta\beta}w^i(t)\frac{1}{C}\sum_{c=1}^{C}\left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\tilde{I}_c(t)^{\frac{1-\theta\beta}{1-\theta}}.$$

The last term in the government's budget constraint is the integral $\int_0^\infty e^{-rt}\Pi(t)dt$. Recall that $\Pi(t)$ is defined as the aggregate flow of income net of costs for all firms in the economy in period t. For any individual firm, the discounted present value of this flow is equal to the firm's value at $t=0$. Furthermore, since neither the final or non-tradable good producers make profits, the discounted present value of $\Pi(t)$ starting in $t=0$ must be equal to

$$\int_0^\infty e^{-rt}\Pi(t)dt=F(D,L_0)Q(0).$$

Collecting the results above and using that $G(t)=\bar{G}$ for all t, the government's problem can be reduced to

$$\max_{\{s_c\}_{c=1}^C}\quad\int_0^\infty e^{-\rho t}\left\{(L-L_0)^{\theta\beta}I^{1-\theta\beta}\left(1+\theta\beta\frac{L_0}{L-L_0}\right)\left(\frac{1-\theta}{1-\theta\beta}w^i(t)\right)^{1-\theta}\right\}dt$$

$$\text{s.t.}\quad\int_0^\infty e^{-rt}\left\{w^i(t)\frac{1}{C}\sum_{c=1}^{C}s_c\left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\tilde{I}_c(t)^{\frac{1-\theta\beta}{1-\theta}}+\gamma(\bar{G})\right\}dt=F(D,L_0)Q(0)$$

$$+\frac{(1-\beta)\theta}{1-\theta\beta}\int_0^\infty e^{-rt}w^i(t)\left\{\left(\frac{1}{\alpha_0}\right)^{\frac{\theta}{1-\theta}}\left(\theta\beta I\frac{L_0}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}}+\frac{1}{C}\sum_{c=1}^{C}\left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\tilde{I}_c(t)^{\frac{1-\theta\beta}{1-\theta}}\right\}dt.$$

This expression can be further simplified by using the expression for $\tilde{\bar{I}}_c$, which gives

$$\frac{1}{C}\sum_{c=1}^{C} k_c \left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} \tilde{I}_c(t)^{\frac{1-\theta\beta}{1-\theta}} = \frac{1}{C}\sum_{c=1}^{C} k_c \left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} \tilde{\bar{I}}_c^{\frac{1-\theta\beta}{1-\theta}} \times \frac{Z_c(t)^{\frac{1-\theta\beta}{\Theta}}}{e^{\frac{1-\theta\beta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}$$

where k_c equals s_c for the first term in the budget constraint and equals 1 in the last. Given that $C \to \infty$, we can once again apply the law of large numbers to find

$$\frac{1}{C}\sum_{c=1}^{C} k_c \left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} \tilde{I}_c(t)^{\frac{1-\theta\beta}{1-\theta}} = e^{\frac{1-\theta\beta}{\Theta}\frac{(1-\beta)\theta}{\Theta}\frac{\sigma^2}{4\phi}} \frac{1}{C}\sum_{c=1}^{C} k_c \left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} \tilde{\bar{I}}_c^{\frac{1-\theta\beta}{1-\theta}}.$$

For convenience, define the public good production cost as $\gamma(\bar{G}) = \bar{\pi}\bar{G}Q(t)$ and let

$$\bar{w}^i = \frac{1}{\bar{\pi}}\frac{w^i(t)}{Q(t)} = \frac{1}{I^{\frac{\Theta}{1-\theta}}}\psi^{\psi}\left(\frac{1-\psi}{\kappa}\right)^{1-\psi}\frac{(1+\lambda)L_0}{r+D} e^{\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}} \left\{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}\right\}^{\frac{\Theta}{1-\theta}}.$$

Note that $\bar{w}^i$ does not vary with time, which reduces the government's problem to

$$\max_{\{s_c\}_{c=1}^{C}} \quad (L-L_0)^{\theta\beta}\left(1+\theta\beta\frac{L_0}{L-L_0}\right)\left(\frac{1-\theta}{1-\theta\beta}\bar{w}^i\right)^{1-\theta}\int_0^{\infty} e^{-\rho t}Q(t)^{1-\theta}dt$$

$$\text{s.t.} \quad e^{\frac{1-\theta\beta}{\Theta}\frac{(1-\beta)\theta}{\Theta}\frac{\sigma^2}{4\phi}}\frac{1}{C}\sum_{c=1}^{C} s_c\left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\tilde{\bar{I}}_c^{\frac{1-\theta\beta}{1-\theta}} + \frac{\bar{G}}{\bar{w}^i} = \frac{L_0}{\bar{w}^i(r+D)}\frac{Q(0)}{\int_0^{\infty} e^{-rt}Q(t)dt}$$

$$+\frac{(1-\beta)\theta}{1-\theta\beta}\left\{\left(\frac{1}{\alpha_0}\right)^{\frac{\theta}{1-\theta}}\left(\theta\beta I\frac{L_0}{L-L_0}\right)^{\frac{1-\theta\beta}{1-\theta}} + e^{\frac{1-\theta\beta}{\Theta}\frac{(1-\beta)\theta}{\Theta}\frac{\sigma^2}{4\phi}}\frac{1}{C}\sum_{c=1}^{C}\left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\tilde{\bar{I}}_c^{\frac{1-\theta\beta}{1-\theta}}\right\}.$$

Since $Q(t)$ grows at a constant rate g, we can write $Q(t) = Q(0)e^{gt}$ for a given initial condition $Q(0)$. Assuming that $\rho > (1-\theta)g$ (otherwise the welfare function is not well defined),

$$\int_0^{\infty} e^{-\rho t}Q(t)^{1-\theta}dt = Q(0)^{1-\theta}\int_0^{\infty} e^{((1-\theta)g-\rho)t}dt = \frac{Q(0)^{1-\theta}}{r-(1-\theta)g}.$$

After plugging in this result into the government's objective function, it becomes

$$\max_{\{s_c\}_{c=1}^{C}} \quad (L-L_0(\mathbf{s}))^{\theta\beta}\left(1+\theta\beta\frac{L_0(\mathbf{s})}{L-L_0(\mathbf{s})}\right)\frac{\bar{w}^i(\mathbf{s})^{1-\theta}}{\rho-(1-\theta)\lambda D(\mathbf{s})}$$

$$\text{s.t.} \quad e^{\frac{1-\theta\beta}{\Theta}\frac{(1-\beta)\theta}{\Theta}\frac{\sigma^2}{4\phi}}\frac{1}{C}\sum_{c=1}^{C} s_c\left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\tilde{\bar{I}}_c(\mathbf{s})^{\frac{1-\theta\beta}{1-\theta}} + \frac{\bar{G}}{\bar{w}^i(\mathbf{s})} = \frac{r-\lambda D(\mathbf{s})}{r+D(\mathbf{s})}\frac{L_0(\mathbf{s})}{\bar{w}^i(\mathbf{s})} + \frac{(1-\beta)\theta}{1-\theta\beta}\times$$

$$\left[\left(\frac{1}{\alpha_0}\right)^{\frac{\theta}{1-\theta}}\left(\theta\beta I\frac{L_0(\mathbf{s})}{L-L_0(\mathbf{s})}\right)^{\frac{1-\theta\beta}{1-\theta}} + e^{\frac{1-\theta\beta}{\Theta}\frac{(1-\beta)\theta}{\Theta}\frac{\sigma^2}{4\phi}}\frac{1}{C}\sum_{c=1}^{C}\left(\frac{1}{\alpha_c}\right)^{\frac{\theta}{1-\theta}}\tilde{\bar{I}}_c(\mathbf{s})^{\frac{1-\theta\beta}{1-\theta}}\right], \tag{E.1}$$

F Sensitivity Analyzes

The different specifications used to estimate the elasticities of agglomeration and congestion in section 4 produced different point estimates, sometimes significantly different from one another. To account for this variation, we redo step three of the model's estimation procedure (section 4.3) and re-run the counterfactual experiments (section 5) using different values for these elasticities. To span the full range of estimated elasticities, we consider all possible combinations of $\eta \in \{0.15, 0.20, 0.25\}$ and $\beta \in \{0.5, 0.6, 0.8\}$. The counterfactual results are shown in table F.9, which fixes the subsidy cap at 50%.

Note from table F.9 that, despite the wide variation in the values of those two parameters, the optimal subsidies all follow the same pattern: they increase the spatial concentration of the population, increasing the rate of creative destruction/growth, but decreasing the baseline wage of workers. Intuitively, higher values of η and β increase the gains from spatially concentrating the population: a high η means a high elasticity of agglomeration and a high β means a low elasticity of congestion.

Table F.9: Gains from adopting optimal subsidies, $\tau = 0.5$.

		City-level Subsidies			
η	β	Δ Welfare	Δ Baseline Wage	Δ Creative Destruction	Δ Rate of Growth
0.15	0.5	3.97%	−7.29%	1.33 p.p.	0.18 p.p.
0.15	0.6	5.45%	−7.13%	1.80 p.p.	0.24 p.p.
0.15	0.8	15.23%	−6.56%	3.53 p.p.	0.47 p.p.
0.20	0.5	4.12%	−7.20%	1.54 p.p.	0.20 p.p.
0.20	0.6	6.15%	−7.75%	2.00 p.p.	0.26 p.p.
0.20	0.8	18.42%	−7.05%	4.09 p.p.	0.54 p.p.
0.25	0.5	4.61%	−7.60%	1.68 p.p.	0.22 p.p.
0.25	0.6	6.97%	−8.05%	2.18 p.p.	0.29 p.p.
0.25	0.8	22.58%	−7.59%	4.76 p.p.	0.63 p.p.

		State-level Subsidies			
η	β	Δ Welfare	Δ Baseline Wage	Δ Creative Destruction	Δ Rate of Growth
0.15	0.5	1.91%	−4.34%	0.82 p.p.	0.11 p.p.
0.15	0.6	2.86%	−4.48%	1.03 p.p.	0.14 p.p.
0.15	0.8	8.67%	−5.01%	2.24 p.p.	0.30 p.p.
0.20	0.5	2.14%	−4.50%	0.88 p.p.	0.12 p.p.
0.20	0.6	3.23%	−4.71%	1.13 p.p.	0.15 p.p.
0.20	0.8	10.68%	−5.83%	2.68 p.p.	0.35 p.p.
0.25	0.5	2.39%	−4.66%	0.95 p.p.	0.13 p.p.
0.25	0.6	3.65%	−4.83%	1.23 p.p.	0.16 p.p.
0.25	0.8	13.79%	−6.70%	3.30 p.p.	0.44 p.p.

G Extensions of the Model

This section modifies the model in sections 2 and 3 in two important ways. First, we address the issue that the model in the main text predicts that all firms located in the same city, regardless of their size, will have the same arrival rate of innovation. Translating to the data, this means that big and small firms produce the same expected number of patents over any given year, which is course not true. In the following section, we extend the model to allow the production of innovation to scale with firm size in the spirit of Klette and Kortum (2004). Keeping with the interpretation of maintenance and managerial costs, we assume that each firm's fixed costs also scale up with their size.

In this extended model, firms that own more product lines will also hire more inventors and produce more patents, but the innovation *intensity* of firms (patents per worker) will be constant within a city. We will show that the predictions of this extended model concerning the spatial distribution of inventors and innovation and its relationship with economic growth are the exactly the same as the simplified model shown in the main text.

The second modification of the model involves the inclusion of moving costs for inventors. Free mobility of inventors across cities is better understood as reflecting long-run dynamics, while ignoring short-run costs. As an alternative, we consider a model in which the population of inventors in each city is fixed. Under this scenario, the spatial distribution of tax credits has no effect on the rate of growth of the economy, which is counterfactual.

G.1 A Model With Scalable Innovation

We start by redefining the innovation production function for a firm f located in city c as

$$x_{f,c}(t) = \chi_c(t) \left(\tilde{I}_c(t)^{\eta} i_{f,c}(t) \right)^{\psi} p_f(t)^{1-\psi} \tag{G.1}$$

where $p_f(t) = 1 + |\mathbf{q}_f(t)|$ and $\mathbf{q}_f(t)$ is the set of product lines owned by the firm. We define p_f as the number of product lines plus one to maintain symmetry between entrant and incumbent firms (note that entrant firms do not own any product lines, so $\mathbf{q}_f = \varnothing$). We also assume that the fixed cost of innovation scales in the same way, so that firms must hire κp_f inventors to cover their fixed costs of innovation.

All of the remaining assumptions of the model in section 2 are kept, and any results not explicitly shown to be different will still apply to this extension (e.g., lemma 2). The HJB equation for an incumbent is thus given in lemma G.1, which is stated without proof (the proof is exactly analogous to lemma 3.

Lemma G.1. *The HJB equation that describes the problem faced by an incumbent firm located* *in city* $c \in \{1, \ldots, C\}$ *is*

$$rV_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) - \frac{\partial V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)}{\partial A}\frac{\partial A}{\partial t} =$$

$$\max_{x_{f,c}} \left\{ \begin{array}{l} \sum_{q_j \in \mathbf{q}_f} \bar{\pi} L_0 q_j + x_{f,c}\mathbb{E}_j[V_c(\mathbf{q}_f \cup_+ \{(1+\lambda)q_j\}, \tilde{I}_c, Z_c, A) - V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A)] \\ \\ -(1-s_c)w_c^i(i_{f,c} + \kappa p_f) - D \sum_{q_j \in \mathbf{q}_f} [V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) - V_c(\mathbf{q}_f \backslash_- \{q_j\}, \tilde{I}_c, Z_c, A)] + R_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) \end{array} \right\}$$

$$x_{f,c} = \bar{\chi}_c Z_c (\tilde{I}_c^{\eta} i_{f,c})^{\psi} p_f^{1-\psi}$$

The HJB function of an entrant firm will be exactly analogous, with the exception that it own no product lines of its own. Building on lemma G.1, proposition G.1 presents the solution of the firm's problem.

Proposition G.1. *In a Stationary Balanced Growth Path Equilibrium where the total production of final goods* Y *grows at rate* $g < r$*, the value function of an incumbent firm located in city* $c \geq 1$ *and whose portfolio of products is* $\mathbf{q}_f$ *is*

$$V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = F(D, L_0) \sum_{q_j \in \mathbf{q}_f} q_j + \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0) p_f Q\right\},$$

where $F(D, L_0) = \bar{\pi} L_0/(r + D)$ *is the "franchise value" of adding a new product to the portfolio and* E_c *is the entry value for firms city* c *(see the proof for a complete characterization).*

In addition, the second stage value function of an entrant firm who is located in city c *is*

$$V_c^e(\tilde{I}_c, Z_c, A) = \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0) Q\right\}.$$

Proof. The proof of this proposition mimics the proof of proposition 1. We start by guessing the solution

$$V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = F \sum_{q_j \in \mathbf{q}_f} q_j + E_c p_f Q.$$

Substitute this into the firm's HJB to find

$$rF\sum_{q_j\in\mathbf{q}_f} q_j + rE_c p_f Q - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t}p_f Q - E_c p_f \dot{Q} =$$

$$\max_{x_{f,c}}\left\{\begin{array}{l} \bar{\pi}L_0\sum_{q_j\in\mathbf{q}_f} q_j + x_{f,c}[F(1+\lambda)Q + E_c Q] - D\sum_{q_j\in\mathbf{q}_f} Fq_j \\ \\ -(1-s_c)w_c^i\left(\dfrac{x_{f,c}^{\frac{1}{\psi}}}{(\bar{\chi}_c Z_c)^{\frac{1}{\psi}}\tilde{I}_c^{\eta}p_f^{\frac{1-\psi}{\psi}}} + \kappa p_f\right) \\ \\ +\phi\left(\dfrac{\sigma^2}{4\phi} - \ln(Z_c)\right)Z_c\left[\dfrac{\partial E_c}{\partial Z_c} + \dfrac{\partial E_c}{\partial \tilde{I}_c}\dfrac{\partial \tilde{I}_c}{\partial Z_c}\right]p_f Q + \dfrac{(\sigma Z_c)^2}{2}\left[\dfrac{\partial^2 E_c}{\partial Z_c^2}\right]p_f Q \\ \\ +\dfrac{(\sigma Z_c)^2}{2}\left[\dfrac{\partial E_c}{\partial \tilde{I}_c}\dfrac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \dfrac{\partial^2 E_c}{\partial \tilde{I}_c^2}\left(\dfrac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + 2\dfrac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c}\dfrac{\partial \tilde{I}_c}{\partial Z_c}\right]p_f Q \end{array}\right\}$$

The first-order condition is

$$[x_{f,c}]: \quad F(1+\lambda)Q + E_c Q - \frac{1}{\psi}\frac{(1-s_c)w_c^i x_{f,c}^{\frac{1-\psi}{\psi}}}{(\bar{\chi}_c Z_c)^{\frac{1}{\psi}}\tilde{I}_c^{\eta}p_f^{\frac{1-\psi}{\psi}}} \leq 0.$$

If the optimal solution is interior, the arrival rate of an innovation is given by

$$x_{f,c} = p_f\bar{\chi}_c^{\frac{1}{1-\psi}}\left\{\psi[F(1+\lambda)+E_c]\frac{Q}{w^i}\frac{\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c}\tilde{I}_c^{\eta-\frac{(1-\beta)\theta}{1-\theta}}\right\}^{\frac{\psi}{1-\psi}}Z_c^{\frac{1}{1-\psi}}.$$

Plugging $x_{f,c}$ into the HJB equation and collecting terms with and without $\sum_{q_j\in\mathbf{q}_f} q_j$ gives

$$F(D, L_0) = \frac{\bar{\pi}L_0}{r+D}, \tag{G.2}$$

where the same argument made in the proof of proposition 1 applies to show that F is constant over time. In addition, using that $\dot{Q}/Q = g$,

$$
\begin{aligned}
(r-g)E_c - \frac{\partial E_c}{\partial A}\frac{\partial A}{\partial t} &= (1-\psi)\left\{\bar{\chi}_c[F(1+\lambda)+E_c]\right\}^{\frac{1}{1-\psi}} \left[\psi \frac{Q}{w^i}\frac{\alpha_c^{\frac{\theta}{1-\theta}}}{1-s_c}\tilde{I}_c^{\frac{\eta(1-\theta)-(1-\beta)\theta}{1-\theta}}\right]^{\frac{\psi}{1-\psi}} Z_c^{\frac{1}{1-\psi}} \\
&\quad -(1-s_c)\kappa\frac{w^i}{Q}\left(\frac{\tilde{I}_c^{1-\beta}}{\alpha_c}\right)^{\frac{\theta}{1-\theta}} + \phi\left(\frac{\sigma^2}{4\phi} - \ln(Z_c)\right) Z_c \left[\frac{\partial E_c}{\partial Z_c} + \frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right] \\
&\quad + \frac{(\sigma Z_c)^2}{2}\left[\frac{\partial^2 E_c}{\partial Z_c^2} + \frac{\partial E_c}{\partial \tilde{I}_c}\frac{\partial^2 \tilde{I}_c}{\partial Z_c^2} + \frac{\partial^2 E_c}{\partial \tilde{I}_c^2}\left(\frac{\partial \tilde{I}_c}{\partial Z_c}\right)^2 + 2\frac{\partial^2 E_c}{\partial \tilde{I}_c \partial Z_c}\frac{\partial \tilde{I}_c}{\partial Z_c}\right],
\end{aligned}
\tag{G.3}
$$

which implicitly defines $E_c = E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)$.

If the optimal solution is a corner, then both the number of inventors hired and the fixed cost are zero. In this case it is straightforward to show that F, as defined in equation (G.2), and $E_c = 0$ solve the firm's HJB. Summarizing both cases, we can conclude that

$$
V_c(\mathbf{q}_f, \tilde{I}_c, Z_c, A) = F(D, L_0)\sum_{q_j \in \mathbf{q}_f} q_j + \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)Q\right\}.
$$

Entrant's Problem. The entrant's problem is solved in exactly the same way. Repeating the argument made in the proof of proposition 1 shows that

$$
V_c^e(\tilde{I}_c, Z_c, A) = \max\left\{0, E_c(\tilde{I}_c, Z_c, w^i/Q, D, L_0)Q\right\}.
$$

□

We now impose free entry into all cities, which drives the value of entry to zero. Setting $E_c = 0$ in equation (G.3) determines the population of inventors in each city. Proposition G.2 shows how the economy is spatially distributed in this model.

Proposition G.2. *Imposing (1) free entry, (2) labor market clearing for both inventors and production workers, and (3) assuming a large number of cities $C \to \infty$ (so that the Law of Large Number applies and the average of city-specific shocks converges to its mean), the population of inventors in each city is given by*

$$
I_c = I \times \frac{\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}}{\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}} \times \frac{Z_c^{\frac{1-\theta}{\Theta}}}{e^{\frac{1-\theta}{\Theta}\left(\frac{1-\theta}{\Theta}-1\right)\frac{\sigma^2}{4\phi}}}
\tag{G.4}
$$

where $\Theta = (1-\beta)\theta - \psi\eta(1-\theta)$. Moreover, the arrival rate of an innovation for a firm f located in city c is

$$
x_{f,c} = p_f\left(\kappa\frac{\psi}{1-\psi}\right)^{\psi}\bar{\chi}_c \tilde{I}_c^{\psi\eta} Z_c
\tag{G.5}
$$

and the number of inventors hired by each firm in city c is $i_{f,c} = \frac{\psi}{1-\psi}\kappa p_f$. Now let N_c be the *number of firms investing in R&D and located in city c, and J_c be the total number of products produced by those firms. Then,*

$$N_c + J_c = \left(\frac{1-\psi}{\kappa}\right) I_c \tag{G.6}$$

Finally, it can also be shown that the population of production workers in city 0 *is proportional to* L *(i.e.,* L_0 *does not vary over time), and that* w^i *is not affected by the city-specific productivity shocks*

$$\frac{w^i}{Q} \propto \frac{\bar{\pi} L_0}{r+D}\left\{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\bar{\chi}_c}{1-s_c}\right)^{\frac{1-\theta}{\Theta}} \alpha_c^{\frac{\theta}{\Theta}}\right\}^{\frac{\Theta}{1-\theta}}. \tag{G.7}$$

Again, those results are stated without proof since they are exactly analogous to the ones in proposition 2. The only difference in this case is when finding the number of active firms in each city in the equilibrium. This is done by requiring that all inventors living in the city are employed by some firm. Let $\mathcal{F}_c$ be the set of firms f located in city c. Then

$$I_c = \int_{\mathcal{F}_c} (i_{f,c} + \kappa p_f) df,$$

where we integrate over the set of firms f located in city c. Using the expression for $i_{f,c}$ and recalling that $p_f = 1 + |\mathbf{q}_f|$, this expression becomes

$$I_c = \frac{\kappa}{1-\psi}\int_{\mathcal{F}_c} (1 + |\mathbf{q}_f|) df.$$

Let N_c be the number of firms located in city c and J_c the total number products that are produced by those firms. Then

$$I_c = \left(\frac{\kappa}{1-\psi}\right)(N_c + J_c),$$

which gives equation (G.6) after rearranging.

Finally, corollary G.1 determines the aggregate rate of creative destruction in this economy.

Corollary G.1. *The aggregate rate of creative destruction in this economy is*

$$D \propto \frac{1}{C}\sum_{c=1}^{C} \bar{\chi}_c \tilde{I}_c^{1+\psi\eta}, \tag{G.8}$$

Proof. The aggregate rate of creative destruction in this economy is given by

$$\begin{aligned}
D &= \sum_{c=1}^{C} \int_{\mathcal{F}_c} x_{f,c} df \\
&= \left(\kappa \frac{\psi}{1-\psi}\right)^{\psi} \sum_{c=1}^{C} \bar{\chi}_c \tilde{I}_c^{\psi\eta} Z_c \int_{\mathcal{F}_c} p_f df \\
&= \left(\kappa \frac{\psi}{1-\psi}\right)^{\psi} \sum_{c=1}^{C} \bar{\chi}_c \tilde{I}_c^{\psi\eta} Z_c (N_c + J_c) \\
&= \psi^{\psi}\left(\frac{1-\psi}{\kappa}\right)^{1-\psi} \frac{1}{C}\sum_{c=1}^{C} \bar{\chi}_c \tilde{I}_c^{1+\psi\eta} Z_c.
\end{aligned}$$

Note that this is exactly the same expression found in the proof of corollary 1. Since there is no change in how the local shock is defined and the population of inventors is allocated in the same way (proposition G.2), it follows that the rate of creative of creative destruction will also have the same structure as the one shown in corollary 1. □

G.1.1 Estimation

The main challenge with this extended version of the model is estimating its parameters. In particular, equation (G.6) relates the number of firms, products and the population of inventors in each city. Because J_c is included in that expression, we can no longer normalize the measure of products in the economy to 1 without affecting the scale of N_c and I_c.

To determine the relative scale of N_c and J_c, one could solve for the firm size distribution in each city. Let $\mu_c(q,t)$ be the measures of firms with q products located city c in period t. Since firms gain products at rate $x_{f,c}$ and loose them at rate D, it follows that

$$\frac{\partial \mu_c(q,t)}{\partial t} = x_{f(q-1),c}\mu_c(q-1,t) + (q+1)D\mu_c(q+1,t) - x_{f(q),c}\mu_c(q,t) - qD\mu_c(q,t)$$

where $x_{f(q),c}$ is the arrival rate of innovation for a firm with q products. The first and second terms in the RHS of the equality account for the inflow of firms into size q: firms with $q-1$ products who gain a new product line and firms with $q+1$ products who loose one; the third and fourth terms account for the outflow: firms with q product lines who gain one extra product and those who loose one product. Note from this differential equation that the measure of firms who own q products in city c is not stationary, as it depends on the realization of the local sock through the arrival rate $x_{f,c}$. As a result, the average number of products per firm in each city is also going to vary over time, making the relationship between N_c and J_c hard to pin down.

The simplified model in the main text will also have a non-stationary firm size distribution in each city. However, that model does not require the average number of products per firm to be determined before it is taken to the data (since the number of products does not affect the number of inventors hired). In addition, both the simplified and extended models make the same predictions about how R&D subsidies affect the distribution of the population and the aggregate rate of growth of the economy. For those reasons, the model introduced in section 2 is chosen for performing the counterfactual policy exercises in this paper.

G.2 A Model With Exogenous City Populations

When the population of inventors and production workers is fixed, most of the equations in the static portion of the model still hold, with the exception that production workers' and inventors' utilities are no longer equalized across cities. Most of the arguments in Proposition 1 also hold under this new assumption. The main difference appears in the free entry condition. In the baseline model, this condition pins down the number of inventors in each city (see the proof of Proposition 2). Now, since this population is exogenous, it instead pins down wages in each city. The free entry condition is

$$0=(1-\psi)[\bar{\chi}_cF(1+\lambda)]^{\frac{1}{1-\psi}}\left[\psi\frac{Q}{w_c^i}\frac{\tilde{I}_c^\eta}{1-s_c}\right]^{\frac{\psi}{1-\psi}}Z_c^{\frac{1}{1-\psi}}-(1-s_c)\kappa\frac{w_c^i}{Q},$$

which gives

$$\frac{w_c^i}{Q}=\left(\frac{1-\psi}{\kappa}\right)^{1-\psi}\frac{(1+\lambda)F}{1-s_c}\left(\psi\tilde{I}_c^\eta\right)^\psi\bar{\chi}_cZ_c.$$

The remaining steps in solving for the equilibrium are very similar. Plugging this into the rate of innovation (from the FOC of the firm's dynamic problem)

$$x_{f,c}=\bar{\chi}_c^{\frac{1}{1-\psi}}\left\{\psi F(1+\lambda)\frac{Q}{w_c^i}\frac{\tilde{I}_c^\eta}{1-sc}\right\}^{\frac{\psi}{1-\psi}},$$

we find that the rate at which firms innovate is still given by

$$x_{f,c}=\left(\frac{\psi}{1-\psi}\kappa\right)^\psi\bar{\chi}_cZ_c\tilde{I}_c^{\psi\eta},$$

but now with an extremely important difference: since the population of inventors in each city is fixed, local rates of innovation *do not react to local subsidies to R&D*. Instead, all the subsidy does is increase the wages of inventors, undoing any potential gains that firms might have. As a result, the arrival rate of innovation is not affected. From the production function of innovation, the number of inventors hired per firm is still

$$i_{f,c}=\frac{\psi}{1-\psi}\kappa$$

and the number of firms in each city is

$$N_c=\left(\frac{1-\psi}{\kappa}\right)I_c,$$

again unchanged by local subsidies to R&D. As a result, it is clear that neither the aggregate rate of innovation

$$D=\sum_{c=1}^{C}N_cx_{f,c}$$

nor the rate of growth of the economy $g=\lambda D$ are affected by the underlying spatial distribution of R&D subsidies.