NBER WORKING PAPER SERIES

THE GEOGRAPHY OF INNOVATIVE FIRMS

Craig A. Chikis
Benny Kleinman
Marta Prato

Working Paper 34010
http://www.nber.org/papers/w34010

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2025, Revised April 2026

We are grateful to the University of Chicago, including the Becker Friedman Institute for Economics, Stanford University, and Bocconi University, for research support. We are grateful to Emin Dinlersoz, Pablo D. Fajgelbaum, Elisa Giannone, Xian Jiang, Martí Mestieri, and Javier Quintana for discussing the paper. We thank Ufuk Akcigit, Adrien Bilal, Cécile Gaubert, Jonathan Goldberg, Chad Jones, Pete Klenow, David López-Salido, Esteban Rossi-Hansberg, Robert Shimer, Christopher Tonetti, as well as conference and seminar participants at multiple institutions, for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Geography of Innovative Firms
Craig A. Chikis, Benny Kleinman, and Marta Prato
NBER Working Paper No. 34010
July 2025, Revised April 2026
JEL No. E0, F0, L0, O0, R0

ABSTRACT

Firms conducting R&D across multiple local markets account for most U.S. innovation. We study how their geographic scope affects growth and whether it is socially optimal. We develop a spatial growth model with multi-market firms that generate local knowledge spillovers. In equilibrium, firms may expand into too few or too many markets, depending on the sensitivity of spillovers to their local footprint. We estimate the model using data on R&D locations, patents, and citations, and find that firms under-invest in geographic expansion. Through model counterfactuals, we show that welfare gains from policies promoting broader geographic expansion are higher than traditional R&D subsidies.

Craig A. Chikis
The University of Chicago
cachikis@uchicago.edu

Marta Prato
Bocconi University
marta.prato@unibocconi.it

Benny Kleinman
Stanford University
Department of Economics
and NBER
bennyk@stanford.edu

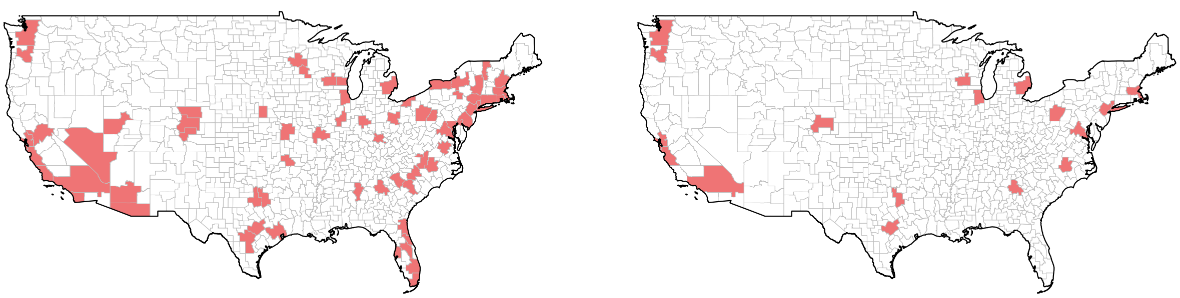
1 Introduction

The majority of U.S. innovation output originates from firms that engage in innovation activity across multiple local markets. Figure 1 illustrates two examples: in 2015, IBM had patenting activity in over 70 local markets, while Google patented in around 20. If such activity generates local knowledge spillovers to surrounding firms and inventors, firms' location decisions affect not only their own innovation output but also that of others. In such an environment, a firm's decision about how many markets to operate in can influence aggregate innovation and economic growth, and may be socially suboptimal.

Figure 1: **Locations with patenting activity of IBM and Google**

(A) IBM

(B) Google



Note: Panel (A) displays a map of IBM's innovation activity across U.S. commuting zones. A commuting zone is counted if IBM has at least one patent with inventors from that location in USPTO data and a recorded establishment in Dun & Bradstreet data. Panel (B) presents the corresponding map for Google. Data from 2015.

In this paper, we study how the geographic structure of innovative firms affects aggregate innovation and growth, whether it is socially optimal, and the welfare implications of policies that alter firms' geographical scope. To address these questions, we develop an endogenous growth model with multiple local markets, multi-location innovative firms, and knowledge spillovers from local firm activity. We characterize the competitive equilibrium and the social planner's solution. We show that firms in equilibrium may operate in too few or too many local markets, depending on the sensitivity of spillovers to their local market footprint. Using data on firms' locations, patents, and citation networks, we provide empirical evidence for the model's key assumptions, including the presence of local knowledge spillovers from firms' innovative locations. Estimating the model, we find parameter values suggesting that U.S. innovative firms operate in too few markets relative to the social optimum. Using the quantified model, we evaluate the effectiveness of policies that promote geographical expansion and show that they outperform conventional R&D subsidies.

Our theoretical framework extends the canonical "creative destruction" model from the economic growth literature to a multi-market setting with two key components. First, we introduce multi-location innovative firms that employ R&D workers across multiple local markets. Firms incur a fixed cost for each location, generating increasing returns that incentivize the concentration of their innovation activity. At the same time, firms face diminishing labor efficiency within each location, generating incentives for firm-level geographical diversification.[1]

Second, following the economic geography literature, we allow the productivity of R&D workers in each market to depend on the mass and composition of local R&D employment and innovative firms. Importantly, we assume that these spillovers exhibit diminishing learning from each individual firm, such that workers in different firms are imperfect substitutes in the regional agglomeration spillovers function. Our assumptions imply that, all else equal, workers in more efficient or larger firms contribute more to local agglomeration spillovers. At the same time, the marginal worker in a given firm may generate smaller spillovers when the firm already employs many local workers. Intuitively, learning from the marginal Google employee in a given market may be limited if Google already has a large local footprint. One way to interpret this spillover structure is that there are gains from a diversity of ideas, as emphasized by Jacobs (1969).[2]

Solving the social planner's problem, we establish that firms in the decentralized equilibrium may operate in either too few or too many markets. On the one hand, when a firm opens an innovation plant in a new market, it generates knowledge spillovers in that location. On the other hand, due to labor-market clearing, new innovation plants reallocate labor from existing ones. If local spillovers from a given plant increase with its size, this reallocation reduces spillovers from the shrinking plants. The optimal number of locations per firm thus depends on how sensitive these spillovers are to the firm's local employment. Intuitively, if knowledge spillovers from a Google office depend only weakly on its size – so that even a single employee generates substantial benefits for nearby firms – a policymaker would prefer Google to open many small offices nationwide. At the opposite extreme, when nearby firms benefit only if Google operates a large local team, the planner might prefer fewer, larger facilities, relative to Google's private choice. In the knife-edge case where the spillover elasticity matches the firm's own innovation elasticity, expansion decisions are efficient even in the presence of knowledge spillovers.[3]

We also examine the spatial allocation of innovation labor, which raises a distinct question from that of firms' optimal spatial scope.[4] When we allow for frictionless mobility of innovation labor, the allocation of innovation labor across space in the baseline model is efficient, despite the presence

[1] A potential micro-foundation for this assumption is that workers are differentiated either by the amenity value they derive from each firm or by the efficiency units they contribute. Firms then face diminishing marginal efficiency of labor in each location.

[2] As in the agglomeration literature, our model implies that regional innovative output scales with the size of the local innovative labor force, so that R&D workers tend to be more productive in larger markets. Our main planning result, however, concerns the optimal allocation of firms across space, taking the allocation of labor as given.

[3] These results resemble optimal entry results in the spirit of Dixit and Stiglitz (1977), although here, the extent of sub-optimal plant entry is determined by the structure of knowledge spillovers across firms rather than the tradeoff between love-of-variety and business stealing.

[4] In particular, even when the spatial distribution of labor is held fixed, the planner may wish to reallocate local resources between idea production (increasing plants' size) and the entry of additional innovative plants within each market (increasing the number of plants).

of knowledge spillovers.[5,6] Notably, a key difference between optimal spatial scope and the spatial allocation of labor lies in the presence of non-rivalries: while labor can be in only one market at a time, multiple markets can simultaneously benefit from the knowledge generated by a single firm. In our model, this takes the form of a firm-level entry cost that is paid only once – regardless of the number of markets in which the firm operates – combined with a local spillover structure that generates uninternalized gains from multi-market activity by the same firm.

To assess whether innovative firms operate in too few or too many locations – and to evaluate the welfare consequences of policies that encourage their expansion – we bring the model to data on firm geography and innovation. Our primary data are patent records and cross-firm patent citations from USPTO's PatentsView, which also identifies the set of locations where each firm conducts innovation activity. While patents offer only a partial measure of innovation activity, they offer two key advantages in our context. First, the data in inventors' locations allows us to analyze innovation activity in firms' different locations. Second, we can utilize cross-firm citations as a direct measure of knowledge flows between firms, which in our theory is linked to firms' location decisions. We supplement these data with establishment locations from Dun & Bradstreet, providing an additional verification of firms' geographic footprints.

We document key empirical facts that support the main assumptions of our model. First, innovative firms that operate in multiple markets account for most of the patenting output in the U.S. economy. Second, when these firms expand in space, we find patterns consistent with local knowledge spillovers, as measured by external patent citations received from other firms in their new location. Using an event-study design, we show that a firm's expansion into a local market is followed by a persistent increase in the external citations it receives from inventors in that market, along with higher patenting activity by other firms in the same local technology class. Third, we show that these spillovers indeed increase with the size of the new R&D facility, but less than one-to-one, in line with our assumption of diminishing returns in the spillovers that a firm generates locally. Moreover, we show that a firm's own local innovation output scales more strongly with its local employment than the spillovers it generates for other firms, providing a first indication that firms under-expand in equilibrium.

We estimate a quantitative version of our model and find evidence that U.S. innovative firms operate in too few markets relative to the social optimum. We identify the elasticity of local knowledge spillovers with respect to a firm's local footprint by examining how external patent citations in a given market respond to the cited firm's local employment. We identify firms' innovation elasticity from the responsiveness of their local patent output to their local employment. We consistently estimate the spillover elasticity to be below the innovation elasticity, implying that – through the lens of our model – firms expand too little across space in equilibrium. Another important takeaway

[5] This result follows from the assumption of constant-elasticity spillovers and the absence of compensating differentials across regions, such as differences in amenities. See Fajgelbaum and Gaubert (2020), Glaeser and Gottlieb (2008), and Kline and Moretti (2014).

[6] More generally, in our quantified model – where we allow for realistic labor mobility and compensating differentials across space – the spatial allocation of labor is inefficient. However, we find that the welfare gains from reallocating labor across space are small relative to those from changing the spatial scope of firms.

from our estimation concerns the structure of local agglomeration spillovers. We find a significant role for the diversity of local innovating firms in promoting regional innovation productivity, conditional on regional labor inputs. Importantly, our estimation implies that the planner would like firms to be marginally more dispersed across space, while maintaining the observed innovation clusters and without substantial reallocation of innovation labor across regions.

We utilize our estimated model to assess the potential benefits of incentivizing innovative firms to increase their spatial scope. We explore a series of subsidies aimed at promoting geographic expansion and compare them to the standard R&D subsidy featured in the endogenous growth literature. This comparison allows us to evaluate the additional gains from geographical expansion beyond the typical effect of allocating more resources to innovation. We calibrate the size of all programs to 0.10% of GDP, which is approximately the cost of R&D support in the United States. A standard R&D subsidy of this magnitude results in a 0.20% increase in consumption-equivalent (CE) welfare relative to the baseline equilibrium. In contrast, a proportional expansion subsidy – which targets the ratio of expansion to R&D expenditure identified in our theory – yields CE welfare gains of approximately 0.30%, about 50% more than the R&D subsidy. A simpler per-market expansion subsidy, which grants firms a flat transfer for each additional market, generates even larger gains of 0.43% at this policy scale, while also reducing spatial inequality. We also find that, depending on implementation, incentivizing spatial expansion might result in greater misallocation of labor across space. However, we find this effect to be quantitatively very small relative to the gains from increasing firms' spatial scope.

Incentivizing spatial expansion has distinct distributional implications compared to standard R&D subsidies. To illustrate this, we decompose our preferred welfare metric into two components: aggregate real income gains (in present discounted value) and a residual distributional effect that captures inequality across space, as well as inequality between innovation and production workers. All the policies we consider have a negative distributional effect, as they increase income more for ex-ante richer innovation workers. However, the negative distributional effect is less pronounced when incentivizing spatial expansion. While R&D subsidies tend to boost income in high-wage regions like Silicon Valley, exacerbating spatial disparities, the per-market expansion subsidy is the only policy that combines aggregate gains with spatial convergence, as it generates a greater incentive to expand into less profitable markets. We also find that at larger policy scales, the proportional expansion subsidy dominates, while the per-market subsidy overshoots the planner's optimal allocation by pushing low-productivity firms to expand into markets where the spillover benefits do not justify the resource cost.

Related literature. This paper relates to several strands of literature.

First, we contribute to the theoretical literature on endogenous growth, combining elements of both innovation-based and diffusion-based growth models. On the innovation side, we present a quality-ladder framework, building on creative destruction models following Grossman and Helpman (1991), Aghion and Howitt (1992), Klette and Kortum (2004), and Akcigit and Kerr (2018). Relative to this literature, we allow innovative firms to operate across multiple local markets, and highlight the importance of these decisions for aggregate growth in the presence of local spillovers. On the knowledge diffusion side, we relate to models of idea flows, in which agents improve their innovation process by learning from others, such as Lucas and Moll (2014), Perla and Tonetti (2014), Akcigit et al. (2018), Buera and Oberfield (2020), and Prato (2025). This literature has also examined the role of knowledge diffusion in space as a source of long-run growth (Berkes et al., 2025; Cai et al., 2025; Eckert and Peters, 2025), and the role of trade in diffusing knowledge across locations (Sampson, 2015; Santacreu, 2015; Perla et al., 2021; Cai et al., 2022; Ayerst et al., 2023; Hsieh et al., 2023; Lind and Ramondo, 2023). Relative to this literature, we connect knowledge diffusion to location decisions by multi-plant firms, in an environment that features local knowledge spillovers. We highlight that in such a setting, the amount of knowledge diffusion in the economy depends on firms' geographical scope, and both innovation and expansion decisions can be sub-optimal in equilibrium.

Second, we relate to the literature on multi-location firms in spatial economics, including Argente et al. (2020), Kerr (2020), Kleinman (2023), and Oberfield et al. (2024); and the related literature on multinational firms in international trade, including Helpman (1984), Garetto (2013), Keller and Yeaple (2013), Ramondo and Rodríguez-Clare (2013), Tintelnot (2017), and Arkolakis et al. (2018). Relative to these two strands of literature, we emphasize that firms' geographic scope may be suboptimal from a planner's perspective when local plants generate local knowledge spillovers. In addition, our analysis focuses on the spatial distribution of firms' innovation activities, rather than on their production structure.

Third, we relate to the literature that studies the optimal allocation of labor across space, e.g. Glaeser and Gottlieb (2008), Kline and Moretti (2014), Fajgelbaum and Gaubert (2020), Fajgelbaum and Gaubert (2025), and Rossi-Hansberg et al. (2025). In contrast to this literature, we focus on the optimal allocation of resources to firm expansion, which in our model takes the form of the allocation of labor between fixed and variable costs in each location. We highlight that for a given allocation of labor across space, firms can be too spatially dispersed or concentrated. We also relate to recent papers that study the role of local spillovers between workers for aggregate productivity and growth, e.g. Martellini (2022), Crews (2023), and Lhuillier (2023). Relative to these papers, we focus on knowledge spillovers between firms, highlighting a new connection between the geographical scope of firms and aggregate growth and deriving new policy implications for the optimal geography of firms.

Empirically, we relate to the literature that investigates the local nature of spillovers between firms, documenting the role of technology clusters (Jaffe et al., 1993; Audretsch and Feldman, 1996; Peri, 2005), and in particular the existence of spillovers on firms' productivity (Greenstone et al., 2010; Bloom et al., 2013; Giroud et al., 2024), and spillovers on innovation activity (Griffith et al., 2011; Carlino and Kerr, 2015; Berkes and Gaetani, 2020; Moretti, 2021; Matray, 2021; Pauly and Stipanicic, 2025; Giroud et al., 2026); see Chatterji et al. (2014) and Kerr and Robert-Nicoud (2020) for a review. Our main empirical design has two advantages relative to these papers. First, by studying changes in patent citations following firm expansion, we provide direct, well-identified evidence for knowledge spillovers that result from changes in firms' geographical scope. Second, our large-scale data and matching design allows us to investigate heterogeneity in the effect of different plant expansions. In particular, as our theory shows, understanding how knowledge spillovers relate to the size of the expanding plant is crucial to determining whether there is too much or too little geographical firm expansion in equilibrium. Finally, it is worth noting that throughout the paper, we focus on spillovers in innovation, and not on contemporaneous productivity spillovers.

The rest of the paper is organized as follows. Section 2 introduces the theoretical model of endogenous growth with multi-location innovative firms and local knowledge spillovers, and characterizes the decentralized equilibrium and the social planner's solution. Section 3 presents empirical evidence supporting the model's key assumptions. In Section 4 we review the estimation of the model for the U.S. economy. In Section 5 we use the estimated model to study the welfare and distributional implications of alternative R&D and spatial-expansion policies. Section 6 concludes.

2 A theory of innovative firms with local spillovers

In this section, we introduce a model of innovation-driven endogenous growth with multi-location innovative firms and local knowledge spillovers. Growth is fueled by quality-enhancing innovations produced by firms that hire R&D workers, subject to decreasing returns to R&D labor within each local market. The productivity of R&D workers depends on local agglomeration spillovers, which are shaped by the equilibrium composition of innovative firms that establish a presence in each location. Firms internalize the private benefits of local knowledge when opening additional establishments but do not internalize the spillovers they generate for other local firms.

2.1 Environment

2.1.1 Setting

Time is continuous and indexed by t. The economy consists of N locations, indexed by n. Each location $n \in \{1, ..., N\}$ is populated by a measure $\bar{L}_n$ of households that can supply labor to the production sector and a measure $\bar{H}_n$ of households that can supply labor to the innovation sector. For now, we abstract from labor mobility across space and across sectors, to highlight more clearly the new insights from our framework, but we relax this assumption later in Section 2.4, and in the quantitative model in Section 4. All households discount the future at rate ρ and are endowed with one unit of labor. Households consume a freely tradable final good. The preferences of a household in location n, sector $s \in \{p, i\}$ (where p indicates production and i innovation), are represented by the following utility function:

$$U_{ns0} = \int_0^\infty e^{-\rho t} \frac{(C_{nst})^{1-\phi} - 1}{1 - \phi} \, dt,$$

where C_{nst} is the amount of final-good consumption in period t and ϕ is a parameter that governs the intertemporal elasticity of substitution.

2.1.2 Production technology

The final good Y_t is produced using a continuum of intermediate goods y_{jt} according to the production function:

$$\log Y_t = \int_0^1 \log(y_{jt})dj. \tag{1}$$

Intermediate goods are produced by production labor using the following production function:

$$y_{jt} = A_{jt}\left(\sum_{n=1}^{N}(Z_n L_{jnt})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{2}$$

where A_{jt} is the state-of-the-art technology in the production of good j, L_{jnt} is the mass of production workers from location n that participate in the production of j, Z_n is the relative productivity of production workers in n, and σ is the elasticity of substitution between production workers from different locations.

Growth in this economy occurs due to improvements in A_{jt}, keeping relative productivity across locations constant. The time-invariant local productivity component Z_n and the imperfect substitution across production workers from different locations allow for a realistic degree of heterogeneity in local employment and wages, but are not a key component of the analysis. In this framework, we intentionally keep the structure of the production sector simple and focus primarily on the organization of firms within the innovation sector. Specifically, we abstract from the question of firm boundaries and firms' spatial organization in the production sector.

2.1.3 Innovation

Growth results from improvements in intermediate technology A_{jt} due to investment in innovation, which produces new ideas. Innovation is undirected, and each new idea increases A_{jt} for some intermediate $j \in [0,1]$ by a factor of γ. Ideas are produced by innovative firms, with endogenous mass M_t and indexed by ω, which hire innovation-sector labor across different markets to produce ideas.

The arrival rate of ideas for a firm ω, $\lambda_{\omega t}$, depends on its employment of R&D labor, and is given by

$$\lambda_{\omega t} = \lambda_0 \sum_{n=1}^{N} K_{nt} z_\omega (\ell_{n\omega t})^\eta, \tag{3}$$

where z_ω is the innovation productivity of firm ω; $\ell_{n\omega t}$ is the mass of inventors that firm ω hires in location n at time t; K_{nt} is the endogenous relative productivity of inventors in market n, which depends also on the activity of other firms in that area through local knowledge spillovers; η captures decreasing returns from employing more labor in location n; and λ_0 is a scaling constant.

The assumption of decreasing returns at the market level can be micro-founded as firms facing an upward-sloping supply of labor in each location – for instance, due to heterogeneity in workers' effective labor supply across firms – which generates diminishing labor efficiency in each local market.[7] An alternative interpretation is that firms distribute workers across markets to access local knowledge, but face diminishing returns to knowledge access within each market.[8] The assumption of $\eta < 1$, which we later verify in the data, is the reason that firms choose to spread their activity across space, despite their output (ideas) being completely tradable and non-rival.

We also assume that setting up an innovation plant in location n is subject to a firm-location-specific fixed cost $f_{n\omega t}$ denominated in units of local innovation labor. These fixed costs discourage firms from fragmenting innovation activity across space, giving rise to increasing returns to scale when local innovation teams are small: innovation output per unit of local labor is non-monotonic – initially increasing with team size and eventually decreasing once local employment becomes sufficiently large.

We impose the distributional assumption that the inverse of $f_{n\omega t}$ is drawn from an i.i.d. Pareto distribution with shape parameter θ and scale f_{max}^{-1}, which makes aggregation particularly tractable. This assumption also implies that a firm's marginal cost of expansion increases with the number of markets in which it operates. In a continuous spatial environment, the distributional assumption is isomorphic to an isoelastic span-of-control cost function in the mass of locations.[9]

The aggregate innovation rate in the economy, $\bar{\lambda}$, is equal to the total innovation efforts across all firms in the economy:

$$\bar{\lambda}_t = \int_{\omega \in \Omega_t} \lambda_{\omega t} d\omega, \tag{4}$$

where Ω_t is the set of innovative firms in the economy.

A new innovative firm can be created by paying a fixed entry cost $f_{e,t}$, denominated in units of the final good.[10] Upon entry, firms draw their idiosyncratic productivity z_ω from a distribution with CDF $\Psi(z)$ and PDF $\psi(z)$. To ensure the existence of a balanced growth path, we impose the technical assumption that the entry cost grows at the same rate as the aggregate economy.

[7] Other micro-foundations include the challenge of hiring and supervising large research teams in the same location, or heterogeneity in how workers value firm-specific amenities.

[8] To see this, rewrite Equation (3) as $\lambda_{\omega t} = \lambda_0 \left[\sum_{n=1}^{N} \left(s_{n\omega t}^{\ell}\right)^{\eta} K_{nt}\right] z_\omega \ell_{\omega t}^{\eta}$, where $\ell_{\omega t}$ denotes the firm's total employment across all markets and $s_{n\omega t}^{\ell} \equiv \frac{\ell_{n\omega t}}{\ell_{\omega t}}$ is the share of employment in market n. The firm's innovative output is then increasing in its total employment (with elasticity η), and in a term that captures access to local knowledge across markets, where the contribution of each market exhibits diminishing returns in the firm's local employment share.

[9] Although we do not model firms' production location decisions explicitly, such considerations are indirectly reflected in the idiosyncratic costs of establishing innovation facilities across locations. Interestingly, empirically, we find that a substantial share of innovation facilities in our data are not colocated with production sites.

[10] Alternatively, we can assume that the entry cost is paid in terms of a special intermediate good, produced one-for-one using the composite of production labor, $\left(\sum_{n=1}^{N} (Z_n L_{jnt})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, or directly in units of production labor. Denominating the entry cost in final goods, production labor, or a production-labor-based composite good allows us to isolate the allocation of resources between firm expansion and R&D within the innovation sector, without conflating it with the broader question of the optimal size of the innovation sector. In any case, the specific choice of denomination for the entry cost does not affect the core insights of our framework.

2.1.4 Spillovers

We adopt a specification for local knowledge spillovers that nests existing approaches in the spatial literature, while allowing both the number and size of local firms to influence agglomeration forces in line with our empirical findings. Specifically, the regional productivity of inventors, K_{nt}, is an increasing function of the mass of firms and inventors that are active in location n, given by

$$K_{nt} = \bar{K}_n \left(\int_{\omega \in \Omega_n} z_\omega \ell_{n\omega t}^{\beta} d\omega \right)^{\alpha}, \tag{5}$$

where Ω_n is the set of firms that open a plant in location n; $\bar{K}_n$ is an exogenous determinant of the relative innovation efficiency of location n; α captures the overall importance of insights from other firms for idea creation; and β captures the degree to which spillovers from a particular firm depend on its presence and employment in market n. One natural interpretation for this structure is that regional innovation productivity scales with the mass of insights that a firm can gain from its peers. In this case, $z_\omega \ell_{n\omega t}^{\beta}$ is the arrival rate of insights from each firm ω, and α is the sensitivity of innovation productivity to insights from others. The arrival rate of insights from each firm ω is increasing with its local employment and with its overall efficiency, with their relative importance determined by β.[11]

To get intuition for the implications of this structure, it is worth considering a few special cases. First, when $\beta = 1$, local innovation productivity is a power-function of total inventors in location n, which resembles standard specifications in the spatial economic literature, with α capturing the strength of agglomeration spillovers. Second, when $\beta = 0$, local innovation productivity is a function of the mass of locally active firms, weighted by their efficiency z_ω. In this case, it is sufficient for a firm ω to open a tiny local plant in n for all other firms in n to fully absorb all available insights from ω. When $\beta = \eta$, local innovation productivity is a function of local innovation output, and the contribution of each firm to local knowledge is proportional to its own innovation output. Finally, when firms are homogeneous in their innovation efficiency z_ω, local knowledge is a Cobb-Douglas function of the mass of locally-active firms and the mass of local inventors.

This completes the set-up of the physical environment as viewed from the perspective of a social planner. Table 1 provides a summary of the preferences, technology, and resource constraints in the economy.

2.2 Equilibrium

We now analyze the equilibrium of the model under perfect competition in final good production and labor markets, and monopoly in the intermediate goods sector. In each intermediate product line j, the producer is a monopolist that owns the rights to the frontier technology A_{jt}. The

[11] In Appendix A.10, we also consider a specification that incorporates spillovers across locations, as well as other extensions of the spillovers function. Our main theoretical results continue to hold under this extension, although these spillovers appear relatively minor in the data compared to local spillovers. See also Berkes et al. (2025) and Comin et al. (2025) for related derivations of local knowledge spillovers.

Table 1: **Summary of the economic environment**

	Preferences
Utility function	$U_{nst} = \int_{h=t}^{\infty} e^{-\rho(h-t)} (C_{nsh}^{1-\phi} - 1)/(1-\phi)\, dh$
	Production
Final good	$Y_t = \exp\left(\int_0^1 \log y_{jt}\, dj\right)$
Intermediate goods	$y_{jt} = A_{jt}\left(\sum_{n=1}^{N} (Z_n L_{jnt})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$
	Innovation
Step size of innovation	γ
Aggregate arrival rate of ideas	$\bar{\lambda}_t = \int_{\omega\in\Omega_t} \lambda_{\omega t}\, d\omega$
Firm-level idea production	$\lambda_{\omega t} = \lambda_0 \sum_{n=1}^{N} K_{nt} z_\omega \ell_{n\omega t}^{\eta}$
Knowledge spillovers	$K_{nt} = \bar{K}_n \left(\int_{\omega\in\Omega_n} z_\omega \ell_{n\omega t}^{\beta}\, d\omega\right)^{\alpha}$
Firm efficiency & fixed costs	$z_\omega \sim \Psi(z)\,, \quad f_{\omega n}^{-1} \sim \text{Pareto}\left(\theta, f_{max}^{-1}\right)$
	Resource constraints
Production labor in n	$\bar{L}_n = \int_0^1 L_{jnt}\, dj$
Innovation labor in n	$\bar{H}_n = \int_{\omega\in\Omega_t} (\ell_{n\omega t} + f_{n\omega t} \times \mathbb{I}\{\ell_{n\omega t} > 0\})\, d\omega$
Output	$Y_t = \sum_{n=1}^{N} \sum_{s\in\{p,i\}} C_{nst} + \int_{\omega\in\Omega_t} f_{et}\, d\omega$

monopolist competes against a fringe of competitive firms that can produce intermediate j with an obsolete technology of quality A_{jt}/γ, where γ is the step-size of technological progress.

We normalize the price of the aggregate good to 1. The price of production labor in market n is denoted by w_{npt}, and that of innovation labor by w_{nit}. The price of intermediate good j is given by p_{jt}. Profits from holding the production rights to product line j are denoted π_{jt}, and the value of a new idea by V_{jt}.

To determine how profits are allocated, we assume that innovation rents are distributed by a national fund to workers in proportion to their wages. Agents' income is then equal to their wage multiplied by a constant ς, where the latter is given by the ratio of aggregate income to wages in the economy. To pin down an interest rate in the economy, we allow households to save in a bond in zero net supply. The standard household maximization problem yields $\phi g = r - \rho$ along the balanced growth path, where r is the interest rate, which is constant along a BGP.[12]

[12] Alternatively, we could assume that households can save by investing in a share of a balanced portfolio of all firms in the economy, B_{nst}. The household's budget constraint would imply that $C_{nst} + \dot{B}_{nst} = w_{nst} + r_t B_{nst}$. This assumption would also yield a standard Euler equation along a BGP, $\phi g = r - \rho$. As another possibility, we could assume that all agents are hand-to-mouth, and allocate profits to an absentee capitalist that spends all income on the final consumption good. In this case, the interest rate r would be exogenous, but all the main results in the paper would still hold.

2.2.1 Production sector solutions

The maximization problem of the final goods producer generates the demand for a particular intermediate j, given by $y_{jt} = \frac{Y_t}{p_{jt}}$. The cost minimization problem for an intermediate monopolist implies that the marginal cost of producing an additional unit of the intermediate j when using the state-of-the-art technology is given by

$$q_{jt} = \frac{1}{A_{jt}} \left(\sum_{n=1}^{N} Z_n^{\sigma-1} \left(w_{npt}\right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}} .$$

The monopolist with state-of-the-art technology engages in limit pricing and charges a price equal to the marginal cost of the competitive fringe, such that the price of the intermediate j is $p_{jt} = \gamma q_{jt}$. As a result, the monopolist's profits are given by $\pi_{jt} = \bar{\pi} Y_t$, where we define $\bar{\pi} \equiv \frac{\gamma-1}{\gamma}$. We note that profits are equal across all intermediate product lines j, so we denote $\pi_{jt} = \pi_t$.

The value of holding the rights to state-of-the-art technology in product line j, V_{jt}, satisfies

$$r_t V_{jt} - \dot{V}_t = \pi_{jt} - \bar{\lambda} V_t.$$

It equals flow profits, π_{jt}, minus the risk of displacement due to a new innovation at rate $\bar{\lambda}$, captured by the last term. Along a balanced growth path, the value V_{jt} satisfies

$$V_{jt} = V_t = \frac{\bar{\pi}}{r - g + \bar{\lambda}} Y_t. \tag{6}$$

2.2.2 Innovation sector solutions

Upon a successful innovation, an innovative firm develops a new frontier technology in a random product line j, displacing the existing incumbent monopolist. The innovative firm sells the technology to an intermediate goods producer, extracting all the surplus. In equilibrium, the price of a new idea – when sold to a producer – is equal to the value of holding the rights to the idea, V_t.[13]

Given the value of a new idea, we can express the problem of an innovative firm, which maximizes the expected value from new innovations net of labor costs in all of its locations:

[13] We assume that an innovative firm sells the technology to a randomly selected intermediate goods producer. The identity of the buyer and the specific market structure for ideas are irrelevant to the main results. Since the value of an idea in Equation (6) does not depend on A_{jt}, the return to innovation is identical across potential buyers, as each gains the market and displaces the incumbent, regardless of the current level of productivity in the intermediate. Only the current incumbent values the innovation less due to the cannibalization of existing profits. However, given the continuum of intermediates, a random match with the incumbent occurs with zero probability.

$$\max_{\{\ell_{n\omega t},\mathbb{I}_{n\omega t}\}_{n=1}^{N}} \lambda_{\omega t}V_t - \sum_{n=1}^{N}\mathbb{I}_{n\omega t}\times\left[w_{nit}\left(\ell_{n\omega t}+f_{n\omega t}\right)\right] \tag{7}$$
$$\text{s.t.} \quad \lambda_{\omega t} = \lambda_0\sum_{n=1}^{N}K_{nt}z_{\omega}(\mathbb{I}_{n\omega t}\ell_{n\omega t})^{\eta},$$

where $\mathbb{I}_{n\omega t}$ is an indicator function for positive activity in location n at time t.

Solving this problem yields the following characterization of firm-level decisions.

Lemma 1. ***Firm-level decisions.***
The solution to problem (7) for a firm with productivity z in market n at time t is characterized by:

(i) Optimal R&D employment, conditional on activity in n:

$$\ell_{nt}(z) = \left(z\frac{\eta K_{nt}\lambda_0 V_t}{w_{nit}}\right)^{\frac{1}{1-\eta}};$$

(ii) Probability of being active in n:

$$\chi_{nt}(z) = f_{max}^{-\theta}\left(\frac{1-\eta}{\eta}\ell_{nt}(z)\right)^{\theta};$$

(iii) Expected fixed cost payment, conditional on activity in n:

$$\bar{f}_{nt}(z) = \frac{\theta}{\theta+1}\left(\frac{1-\eta}{\eta}\ell_{nt}(z)\right).$$

All three objects are increasing in regional innovative productivity K_{nt} and firm productivity z, and decreasing in local wages w_{nit}.[14]

Proof of Lemma 1. See Appendix A.

Finally, a free entry condition allows us to pin down the mass of innovative firms in the economy:

$$f_{e,t} = \int_z\sum_{n=1}^{N}\chi_{nt}(z)\left[\lambda_0 K_{nt}z\ell_{nt}(z)^{\eta}V_t - w_{nit}\left(\ell_{nt}(z)+\bar{f}_{nt}(z)\right)\right]d\Psi(z). \tag{8}$$

In equilibrium, the cost of entry for a new innovative firm is equal to the expected profits from creating a new firm. A summary of all the equilibrium conditions can be found in Appendix A.

We now characterize the relationship between firm expansion and R&D in equilibrium. The following proposition describes the allocation of innovative labor between R&D (variable costs,

[14] Note that optimal employment conditional on activity is the same for all firms with productivity z, so we suppress the index ω and write $\ell_{nt}(z) = \ell_{n\omega t}$ for all ω such that $z_\omega = z$. Notably, this structure also implies that average firm efficiency, conditional on activity, is equalized across markets, a pattern strongly supported by the data.

captured by $\ell_{nt}(z)$ above) and firm expansion (fixed costs). We highlight this result here because, as we show in Section 2.3, the planner's optimal allocation differs from the equilibrium one in the presence of knowledge spillovers.

Proposition 2. ***Allocation to spatial expansion in equilibrium.***
Let $x_{nt}(z) \equiv \chi_{nt}(z)\bar{f}_{nt}(z)$ and $y_{nt}(z) \equiv \chi_{nt}(z)\ell_{nt}(z)$ denote the expected amount of region-n innovation labor allocated to expansion (paying fixed plant costs) and R&D activities (hiring inventors) for type-z firms, respectively. In equilibrium, their ratio is given by:

$$\frac{x_{nt}(z)}{y_{nt}(z)} = \frac{1-\eta}{\eta}\frac{\theta}{\theta+1}.$$

Proof of Proposition 2. See Appendix A.

The allocation of innovation labor between expansion and R&D activities for firms of type z in market n is decreasing in η, increasing in θ, and is independent of firm type. When innovative firms choose whether to expand into a new market, they balance two forces. On the one hand, the benefit of spreading out their R&D workers is higher when η is lower, indicating stronger decreasing returns to labor in each location in the R&D production function. On the other, spreading out R&D workers is associated with the fixed cost of running an additional establishment. The cost of expansion is summarized by θ, the shape parameter of the fixed costs distribution: a higher θ implies less dispersed fixed costs, so that conditional on entry, a larger share of local labor is devoted to paying fixed costs rather than conducting R&D.

Another important feature of the model is the dependence of a firm's local innovation output on regional scale. In equilibrium, the firm's innovation output can be expressed in terms of its own productivity z, exogenous regional productivity $\bar{K}_{nt}$, and total regional innovation employment H_{nt}, as shown in the following proposition.

Proposition 3. ***Local scale and firm innovation.***
In equilibrium, the innovation output of a firm with productivity z in market n at time t, conditional on activity, is given by:

$$\lambda_{nt}(z) = \tilde{C}_\lambda \bar{K}_{nt} z^{\frac{1}{1-\eta}} H_{nt}^{\frac{\eta}{1+\theta}+\frac{\alpha(\theta+\beta)}{1+\theta}}, \tag{9}$$

where H_{nt} is total innovation employment in market n and $\tilde{C}_\lambda > 0$ depends on structural parameters and moments of the productivity distribution but not on n or z.

Proof of Proposition 3. See Appendix A.

A natural benchmark for this expression is the agglomeration literature, in which the output of firms – whether innovation or production output – is modeled as a function of regional scale, such as total regional employment. In that literature, the elasticity of firm output with respect to regional scale is typically governed by a single parameter capturing the strength of agglomeration spillovers. In our setting, this elasticity is $\frac{\eta}{1+\theta} + \frac{\alpha(\theta+\beta)}{1+\theta}$, which decomposes into two components. The first term, $\frac{\eta}{1+\theta}$, is a direct scale effect: an increase in H_{nt} raises average firm size with elasticity $\frac{1}{1+\theta}$, and firm innovation scales with employment at rate η. The second term, $\frac{\alpha(\theta+\beta)}{1+\theta}$, operates through knowledge spillovers: an increase in H_{nt} raises local R&D productivity K_{nt} both by attracting more firms (with the mass of local firms scaling as $H_{nt}^{\frac{\theta}{1+\theta}}$) and by increasing the size of existing firms (with each firm's contribution to spillovers scaling with its local employment at rate β). In the limiting case of $\theta \to \infty$, in which all firms are active in all markets, the elasticity reduces to α, as in the standard agglomeration model.

Under immobile labor, $H_{nt} = \bar{H}_{nt}$. We relax this assumption in Section 2.4 and in the quantitative model. Equation (9) also serves as a key moment for estimating the structural parameters in Section 4.

2.2.3 Regional innovation and aggregate growth

We now focus on a balanced growth path (BGP) along which aggregate productivity, defined by $\log \mathcal{A}_t = \int_0^1 \log A_{jt} dj$, consumption, all components of income, and the value of new innovations, all grow at a constant rate g. The spatial distribution of these objects is invariant along the BGP. We omit time subscripts for the remainder of the analysis.

Along a BGP, aggregate growth depends on total innovation output $\bar{\lambda} = \sum_n \lambda_n$, where $\lambda_n \equiv \int_{\omega \in \Omega_n} \lambda_{n\omega}\, d\omega$ aggregates the innovation output of all firms active in location n. This object is shaped by the geographic distribution of innovative activity. The following proposition characterizes regional innovation output and the resulting growth rate.

Proposition 4. ***Regional innovation output and growth.***
Aggregate output grows at a constant rate g, given by

$$g = \bar{\lambda} \log(\gamma) \tag{10}$$

where $\bar{\lambda} = \sum_{n=1}^{N} \lambda_n$ is the aggregate rate of creative destruction. Regional innovation output is given by

$$\lambda_n = C_\lambda \bar{K}_n M_n^{(1-\eta)+\alpha(1-\beta)} \bar{H}_n^{\eta+\alpha\beta}, \tag{11}$$

where $M_n \equiv \int_z M\chi_n(z)\, d\Psi(z)$ is the regional mass of R&D firms, and $C_\lambda \equiv \bar{z}_\eta\, \bar{z}_\beta^\alpha \left(\frac{\eta(1+\theta)}{\eta+\theta}\right)^{\eta+\alpha\beta}$, with $\bar{z}_\eta$ and $\bar{z}_\beta$ defined as weighted averages of firm productivity:

$$\bar{z}_a \equiv \int_z z \left(\frac{z^{\frac{\theta}{1-\eta}}}{\int_z z^{\frac{\theta}{1-\eta}} d\Psi(z)}\right)^{1-a} \left(\frac{z^{\frac{1+\theta}{1-\eta}}}{\int_z z^{\frac{1+\theta}{1-\eta}} d\Psi(z)}\right)^{a} d\Psi(z), \quad a \in \{\eta, \beta\}.$$

Proof of Proposition 4. See Appendix A.

Equation (11) illustrates that regional innovation output depends on two margins: the mass of locally active firms, M_n, and the mass of local innovation workers, $\bar{H}_n$. The elasticity of regional innovation with respect to R&D labor is $\eta+\alpha\beta$. When $\alpha = 0$ and there are no knowledge spillovers, this elasticity simply reflects the decreasing returns to labor in innovation production captured by η. When $\alpha > 0$, the elasticity further reflects the degree to which spillovers scale with firm size, captured by β. Regional innovation also depends on the mass of locally active firms, with elasticity $(1-\eta)+\alpha(1-\beta)$. This term combines an efficiency gain from spreading R&D activity across more firms under decreasing returns $(1-\eta)$ with a spillover gain from having more firms contribute to local knowledge $(\alpha(1-\beta))$. Notably, the relative importance of these two margins depends on η and β: when β is low relative to η, regional innovation is more sensitive to the number of firms than to the size of individual firms.

How does an improvement in the ability of innovative firms to expand in space affect growth in equilibrium? We answer this question in the following proposition, by considering a shift in the scale of fixed costs, f_{max}, that lowers them across the board.

Proposition 5. ***Firm expansion and growth.***
The elasticity of the growth rate to the scale of the cost of expansion is given by:

$$\frac{\partial \log g}{\partial \log f_{max}} = \frac{(1-\eta)+\alpha\,(1-\beta)}{1+\theta}\left(\frac{\partial \log M}{\partial \log f_{max}} - \theta\right).$$

A decline in f_{max} raises growth if there are decreasing returns at the plant level ($\eta < 1$) or spillovers with diminishing learning ($\alpha > 0$ and $\beta < 1$). The effect holds for a given mass of firms (M), and is amplified under free entry $\left(\frac{\partial \log M}{\partial \log f_{max}} < 0\right)$.

Proof of Proposition 5. See Appendix A.

A reduction in the fixed costs of establishing innovation plants fosters growth through two channels, both amplified under free entry. First, when $\eta < 1$, lower expansion costs enhance innovation efficiency by enabling firms to distribute their activities across more locations, which is beneficial under decreasing returns. Second, in the presence of knowledge spillovers ($\alpha > 0$) and diminishing learning from each firm ($\beta < 1$), expanding the number of locations per firm amplifies knowledge spillovers, thereby increasing local innovation productivity K_n. Under free entry, these effects are reinforced: greater profitability of innovation encourages more firms to enter the innovation sector, which in turn enhances innovation efficiency when there are decreasing returns in both idea production and the spillover function.

2.3 Social welfare

We now turn to characterize the planner's allocation. The model is characterized by a few different externalities. First, the model features the externalities typical of Schumpeterian models, where innovation could be sub-optimal due to two opposing forces. On the one hand, innovative firms do not internalize their effect on knowledge creation due to raising the level of quality upon which future innovations will be built. On the other hand, they do not internalize that their innovation efforts increase the creative destruction rate, lowering the equilibrium value of holding the frontier technology in a product line and thus discouraging innovation by others, a force known as the "business stealing" effect.

In addition, our framework introduces a novel spatial expansion externality. Conditional on the allocation of resources to innovation, potential inefficiencies arise due to uninternalized knowledge spillovers between innovative firms. Firms fail to account for the fact that establishing a new facility in a given location generates knowledge spillovers that benefit other local firms. Additionally, they do not internalize that by drawing labor away from existing firms when opening a new plant, they diminish the knowledge spillovers from these firms, as spillovers depend positively on a firm's local size ($\beta > 0$).

To study the implications of these forces, we consider a planner that maximizes aggregate welfare, defined as a weighted average of the present discounted value (PDV) of utility across all agents. We restrict the planner to choosing solutions that satisfy balanced growth.[15] The planner's problem is then given by:

$$\max_{\left\{[\ell_n(z),\chi_n(z)]_z\right\}_{n=1}^{N},M} \quad \int_0^\infty e^{-\rho t} \sum_{n=1}^{N} \sum_{s\in\{p,i\}} \nu_{nt}^s \frac{\left(C_{ns0}e^{gt}\right)^{1-\phi}-1}{1-\phi}\,dt \tag{12}$$

$$\text{s.t.} \quad g = \log(\gamma)\lambda_0 M \int_z \sum_{n=1}^{N} \chi_n(z)\, K_n z \ell_n^{\eta}(z)\, d\Psi(z)$$

$$K_n = \bar{K}_n \left(M \int_z \chi_n(z)\, z \ell_n(z)^{\beta}\, d\Psi(z)\right)^{\alpha}$$

$$\bar{H}_n = M \int_z \chi_n(z)\left(\bar{f}_n(z) + \ell_n(z)\right) d\Psi(z)$$

$$\bar{f}_n(z) = \frac{\theta}{\theta+1} f_{max}\left(\chi_n(z)\right)^{\frac{1}{\theta}},$$

where $\{\{\nu_{nt}^s\}_{s\in\{p,i\}}\}_{n=1}^N$ capture the (non-negative) weights that the planner places on agents in location n and sector s at time t.

As is standard in Schumpeterian models, innovation in equilibrium may be too high or too low relative to the planner's solution, depending on the balance between intertemporal knowledge spillovers and business stealing. We focus here on the novel spatial expansion externality, which can distort the allocation of innovation labor between R&D and geographic expansion. The following proposition characterizes the planner's optimal division of innovation labor between expansion (plant-level fixed costs) and R&D (variable costs).

Proposition 6. ***Optimal allocation to spatial expansion.***
Let $x_n(z)$ and $y_n(z)$ denote the amount of region-n innovation labor allocated to expansion and

[15] We allow the planner to freely redistribute ex-post through transfers of the final good, thus abstracting from distributional concerns in our analysis. In Appendix A, we show that under these assumptions, the planner's objective can be expressed as a function of aggregate consumption.

R&D activities of type-z firms, respectively. Their ratio in the planner's allocation is given by:

$$\frac{x_n(z)}{y_n(z)} = \frac{\alpha \frac{s_n^\beta(z)}{s_n^\eta(z)} \left(\frac{1-\beta}{1-\eta}\right) + 1}{\alpha \frac{s_n^\beta(z)}{s_n^\eta(z)} \frac{\beta}{\eta} + 1} \frac{1-\eta}{\eta} \frac{\theta}{\theta+1}$$

where $s_n^\eta(z)$ and $s_n^\beta(z)$ are the weights of type z firms in local innovation output and in local spillovers, respectively:

$$s_n^\eta(z) \equiv \frac{\psi(z)\, z x_n(z)^{\frac{\theta}{\theta+1}(1-\eta)} y_n(z)^\eta}{\int_{z'} \psi(z')\, z' x_n(z')^{\frac{\theta}{\theta+1}(1-\eta)} y_n(z')^\eta\, dz'}, \quad s_n^\beta(z) \equiv \frac{\psi(z)\, z x_n(z)^{\frac{\theta}{\theta+1}(1-\beta)} y_n(z)^\beta}{\int_{z'} \psi(z')\, z' x_n(z')^{\frac{\theta}{\theta+1}(1-\beta)} y_n(z')^\beta\, dz'}.$$

Proof of Proposition 6. See Appendix A.

Proposition 6 illustrates how the planner's relative expenditure on fixed plant costs diverges from that in equilibrium (Proposition 2). First, when there are no knowledge spillovers ($\alpha = 0$), both solutions are identical. Second, the two solutions may still coincide even when spillovers are present ($\alpha > 0$), but only if the elasticity of innovation output with respect to local labor (η) exactly matches the elasticity of spillovers with respect to local labor (β). When β is less than η, firms in equilibrium under-invest in expansion compared to the planner's optimal solution. Conversely, when β is relatively high, firms over-invest in expansion. Intuitively, a very low β suggests that, given a level of z, smaller plants contribute to spillovers as much as larger ones. In such cases, the planner would favor more smaller plants to boost innovation efficiency across the economy. On the other hand, when β is high, spillovers scale strongly with plant size, and the planner would prefer fewer, larger plants. In the special case where $\beta = \eta$, firms' equilibrium decisions align exactly with the planner's optimal strategy.

Interestingly, whether firms in equilibrium over- or under-expand does not depend on the magnitude of spillovers, as captured by α. Instead, the direction of inefficiency depends solely on the relative magnitude of β compared to η. However, the extent of the inefficiency is naturally influenced by the importance of knowledge spillovers for innovative firms, i.e., the size of α. In Section 4, we aim to estimate these parameters for the case of the innovation sector in the U.S. economy.

The above proposition characterizes the allocation of innovation labor for a given firm type z and location n. We now continue to describe the optimal allocation of innovation labor across firm types z for a given location n. Later, we relax the assumption of no mobility and characterize the allocation of innovation labor across locations n as well.

Proposition 7. *Optimal allocation across firm types.*
The mass of local innovation labor that the planner allocates to firms of type z is given by:

$$\frac{\alpha s_n^\beta(z)\left(\beta + (1-\beta)\frac{\theta}{\theta+1}\right) + s_n^\eta(z)\left(\eta + (1-\eta)\frac{\theta}{\theta+1}\right)}{\int_{z'}\left(\alpha s_n^\beta(z')\left(\beta + (1-\beta)\frac{\theta}{\theta+1}\right) + s_n^\eta(z')\left(\eta + (1-\eta)\frac{\theta}{\theta+1}\right)\right) d\Psi(z')} \bar{H}_n.$$

Proof of Proposition 7. See Appendix A.

The proportion of local innovation labor allocated to firms of type z reflects both their contribution to local spillovers, captured by the term $\alpha s_n^{\beta}(z)\left(\beta+(1-\beta)\frac{\theta}{\theta+1}\right)$, and their direct contribution to innovation output, captured by $s_n^{\eta}(z)\left(\eta+(1-\eta)\frac{\theta}{\theta+1}\right)$. When spillovers are stronger (i.e., when α is large), the first term in the numerator becomes more important, and the planner places greater weight on firms that contribute more to spillovers. This echoes the well-known insight from the innovation literature that basic research, i.e., being more spillover-intensive, warrants greater public support than applied research. In our context, this implies that the planner optimally shifts resources toward firms that generate more knowledge spillovers. Notably, when $\eta=\beta$, the allocation of innovation labor across firms is identical to the case without spillovers ($\alpha=0$), in which case the equilibrium allocation is efficient.

2.4 Labor mobility

Up until now we have considered the case in which labor is immobile. We now characterize the allocation of labor across space in the simple case of free labor mobility, focusing on the mobility of labor in the innovation sector.[16] In our full quantified model below, we consider a more realistic environment with limited mobility, as well as mobility between the production and innovation sectors.

We assume that there is an exogenous economy-wide endowment of $\bar{H}$ innovation workers. In equilibrium, free mobility implies the equalization of wages across space. For the central planner, this implies solving also for H_n in the optimization problem (12), subject to the constraint $\sum_{n=1}^{N} H_n=\bar{H}$. The next proposition characterizes the allocation of innovation labor across space.

Proposition 8. *Allocation of innovation labor across space.*
Under free labor mobility, the share of innovation labor in market n, given by $H_n/\bar{H}$, is efficient. In both the decentralized equilibrium and the planner's solution, this share is given by

$$\frac{H_n}{\sum_{n'=1}^{N} H_{n'}}=\frac{\left(\bar{K}_n\right)^{\frac{1}{\frac{1-\eta}{1+\theta}-\alpha\frac{\beta+\theta}{1+\theta}}}}{\sum_{n'=1}^{N}\left(\bar{K}_{n'}\right)^{\frac{1}{\frac{1-\eta}{1+\theta}-\alpha\frac{\beta+\theta}{1+\theta}}}}.$$

Proof of Proposition 8. See Appendix A.

The allocation of labor across space is efficient, despite the presence of knowledge spillovers. This result follows from two key assumptions: constant-elasticity spillovers and the absence of compensating differentials across regions (such as differences in amenities) (see Fajgelbaum and Gaubert, 2020, Glaeser and Gottlieb, 2008, and Kline and Moretti, 2014 for further discussion). In our quantified model, we relax this assumption by allowing for compensating differentials, in which case the spatial allocation of labor is no longer efficient. However, we find that the potential gains

[16] The production sector features no spillovers, and the optimal allocation of labor across space in that sector is efficient. We show this result in Appendix A.9.

from reallocating labor across space are quantitatively small, consistent with findings elsewhere in the literature.

The results on the optimal allocation to spatial expansion (Proposition 6) and across firm types (Proposition 7) hold also under free labor mobility. Notably, in this setting, the planner would like to alter firms' spatial scope without changing the allocation of labor across space, increasing firms' spending on plant-level fixed costs relative to their spending on variable costs across all markets when $\beta < \eta$, reducing that ratio when $\beta > \eta$, and leaving it unchanged when $\beta = \eta$.

2.5 Extensions

In Appendix Section A.10, we explore alternative specifications for the spillover function and the cost of expansion. First, we demonstrate that our main results remain robust when accounting for decreasing returns to the number of firm locations, conditional on the labor employed, by introducing a span-of-control cost for the number of locations, as in Oberfield et al. (2024). In this setup, the curvature of the span-of-control cost replaces the role of the fixed cost dispersion parameter, θ. Second, we extend our analysis to a more general spillover function, allowing firms to learn not only from plants in the same location but also from those in other locations, while incorporating spatial frictions that hinder cross-location knowledge transfer. The main takeaways from our theoretical analysis remain the same with this specification. Third, we generalize the spillover function to allow an arbitrary exponent on firm productivity in the spillover contribution. We show that the optimal allocation to spatial expansion is unchanged: whether firms over- or under-expand depends only on the comparison of β and η, not on how strongly firm productivity weights the spillover function. Fourth, we allow for imperfect substitution across locations in the firm's innovation production function, so that being present in multiple markets is more valuable when each location contributes differentiated inputs. The main insight that when β is sufficiently low the planner would like to encourage more spatial expansion continues to hold.

3 Empirical evidence

This section presents evidence supporting key assumptions and implications of our model and examines the empirical moments that inform its estimation. First, we demonstrate that firms conducting R&D activities across multiple markets account for the majority of U.S. innovation output, and that these firms tend to replicate similar innovation activities across their locations. Second, we provide evidence of local knowledge spillovers from firm expansion, consistent with our assumption that a firm's local innovation efficiency (K_{nt}) depends on the presence of other firms in the same market. Third, we provide evidence that spillovers increase with the size of a firm's local R&D facility, but less than one-to-one, and that they scale more weakly with local firm size than does the firm's own innovation output – a pattern that, through the lens of the model, suggests firms under-expand in equilibrium.

3.1 Data

Firm-level data. Our primary data source is PatentsView, which provides information on the universe of U.S. patents granted after January 1976. PatentsView contains detailed information about inventors (including their names and locations), the firms that hold the patents (through disambiguated assignee data), and the types of technologies the patents protect, through the Cooperative Patent Classification (CPC) system. Importantly, PatentsView is sufficient for most of our analysis: it provides both a measure of innovation output (patents and citations) and a measure of firms' innovation locations, since patent applications disclose the addresses of their inventors. We leverage a key, often overlooked feature of the data: the ability to track the geographic distribution of inventors and patenting activity *within* firms across their different locations. We supplement PatentsView with establishment-level data from Dun & Bradstreet (D&B), which provides detailed information on the spatial organization of firms between 1969 and 2021. PatentsView remains our source for patenting, citations, and inventor locations. We use D&B to cross-validate the geographic presence of patenting firms and to construct the inventor-to-firm links used in the patenting-efficiency and estimation exercises.[17] Notably, when measuring the local innovation activity of a firm, we impose the strict requirement of having both recorded inventors in the Patents data and a recorded establishment in the Dun & Bradstreet data.

Matching patents and firms' geographies. We merge these two datasets following the methodology of Hughes et al. (2021), which attempts to merge patents to Dun and Bradstreet records by exact- and fuzzy-matching on firm name. We refine this approach by localizing the search for these firms to regions in the vicinity of firms' self-reported locations, information disclosed in patent applications. Appendix B.1 reports example patent and Dun & Bradstreet records (Tables B.1 and B.2) together with summary merge-quality statistics (Table B.3). The quality of the merge is high: over the period 1969 to 2021, we match 84% of all nonwithdrawn utility patents granted to at least a single U.S.-based assignee. Appendix B.1 shows that this coverage is stable across application years and commuting zones (Figure B.1) and that the matched sample closely tracks the PatentsView technology distribution (Figure B.2).

The resulting merged dataset allows us to jointly observe firms' innovation output, citation flows, and geographic footprints. Table B.4 presents summary statistics, pooled across years. The average firm in our data operates in 2 markets and generates 7 patents per year, while employing 11 inventors.[18]

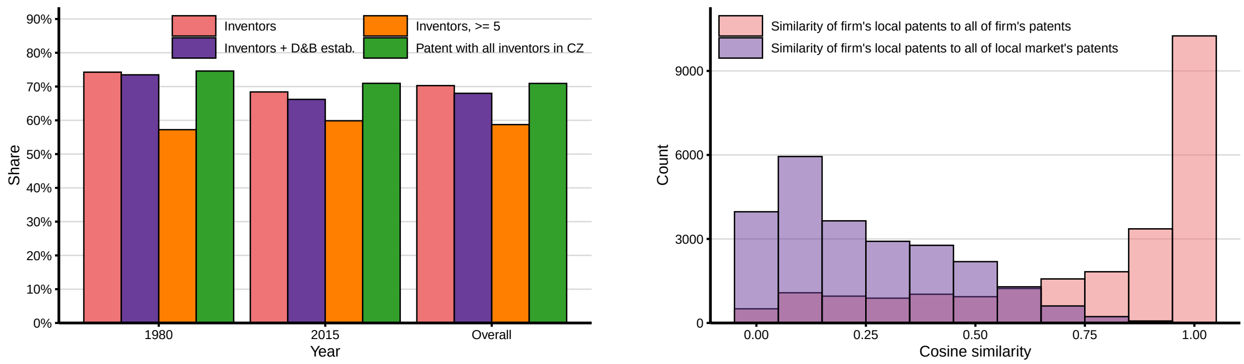
Regional data. Our analysis focuses on spillovers that emerge from firm expansion at the

[17]While D&B data have known limitations for studying establishment dynamics in the full population (Barnatchez et al., 2017), our outcomes are measured from patent records, and D&B enters primarily through cross-validation of establishment presence, the patent-to-firm merge, and the construction of inventor-to-firm links. Moreover, as we show, patenting activity is heavily concentrated among larger, multi-location firms, for which Barnatchez et al. (2017) show that D&B correlates highly with official sources (above 95% for establishment counts and above 80% for employment levels). Appendix B provides additional details.

[18]We also track the co-location of production and innovation activity within firms. Consistent with Fort et al. (2020), we find that co-location has declined over time, aligning with the broader contraction of U.S. manufacturing employment activity: by 2015, only 45% of markets where a firm conducted innovation also had a recorded production establishment.

Figure 2: **Multi-market innovative firms in the U.S.**

(A) Share of total patents by multi-market firms (B) Innovation portfolio similarity across markets



Note: Panel (a) reports the share of total U.S. patents produced by firms that innovate in more than one local market, shown for pooled data and specific years (1980 and 2015). Each bar corresponds to an alternative definition of what constitutes an innovation market for a firm: (i) a commuting zone with at least one inventor listed on the firm's patents; (ii) a commuting zone with at least one inventor and a recorded Dun & Bradstreet establishment; (iii) a commuting zone with at least five inventors; and (iv) a commuting zone with a patent whose inventors all report residence in that commuting zone. Panel (b) shows the distribution of cosine similarity between a firm's local patent portfolio (measured as a vector of patent counts across CPC classes) and two benchmarks: (i) the firm's own portfolio in its other markets (red) and (ii) the average portfolio of other firms in the same local market (blue). Cosine similarity ranges from 0 to 1, with higher values indicating more similar technology mixes. The sample includes firms with patenting activity in multiple markets.

commuting zone level (our notion of a "market"). We obtain county-level information on economic activity and aggregate it to the commuting zone level from various publicly available, tabulated micro-datasets, including the Business Dynamics Statistics and County Business Patterns data of the U.S. Census Bureau.

3.2 The centrality of multi-market innovative firms

We begin by documenting the dominant role of multi-market firms in U.S. innovation. Figure 2A reports the share of total U.S. patents attributed to firms that innovate in more than one local market, under alternative definitions of what constitutes an innovation market for a firm. Our baseline definition classifies a commuting zone as an innovation market for a firm if it contains at least one inventor address listed on the firm's patents. Under this definition, 70% of U.S. patenting output is produced by multi-market innovating firms. We then consider progressively stricter definitions: requiring both an inventor and a D&B establishment in the commuting zone; requiring at least five local inventors; and requiring a patent with all inventors located in that commuting zone. Across these definitions, the share of total patents attributed to multi-market firms ranges from 59% to 71% and is stable across time periods.

Second, we document little differentiation in firms' innovation activities across their multiple markets. We compare the technological similarity of a firm's patent portfolio in each market to (i) the firm's average innovation profile across its other locations and (ii) the average profile of other firms in the same market. High similarity to local peers would indicate market-specific tailoring of innovation (e.g., reflecting local comparative advantage of inventors), whereas high similarity across a firm's own locations would suggest that firms replicate similar innovation activities when expanding to new markets. We measure similarity using the cosine similarity between the firm's local distribution of innovations across knowledge classes and these two benchmarks. Because these vectors contain nonnegative patent counts, cosine similarity ranges from 0 to 1, with higher values indicating more similar technology mixes. Figure 2B shows a histogram for both distances, suggesting that firms' innovations are substantially more similar across their own markets than to those of other firms in the same location.

3.3 Evidence for spillovers from firm expansion

We now provide evidence consistent with the view that changes in firms' spatial scope are accompanied by knowledge spillovers in their new expansion markets, in line with our assumption on the structure of K_{nt}. Following a rich tradition in the empirical and quantitative growth fields (Akcigit and Kerr, 2018; Jaffe et al., 1993), we use external patent citations as proxies for knowledge spillovers. While citations capture only a subset of actual knowledge transfers, they provide a useful measure of citation links between inventors and firms. We focus on two outcome variables: (i) external citations received by the expanding firm from other firms in the same market, which we interpret as suggestive evidence of local knowledge flows; and (ii) patenting efficiency of other local firms in the same technology class, which lets us examine whether these citation patterns are accompanied by broader innovation gains among nearby firms. Together, these outcomes allow us to assess whether firm expansion is associated with patterns consistent with a local spillover channel.

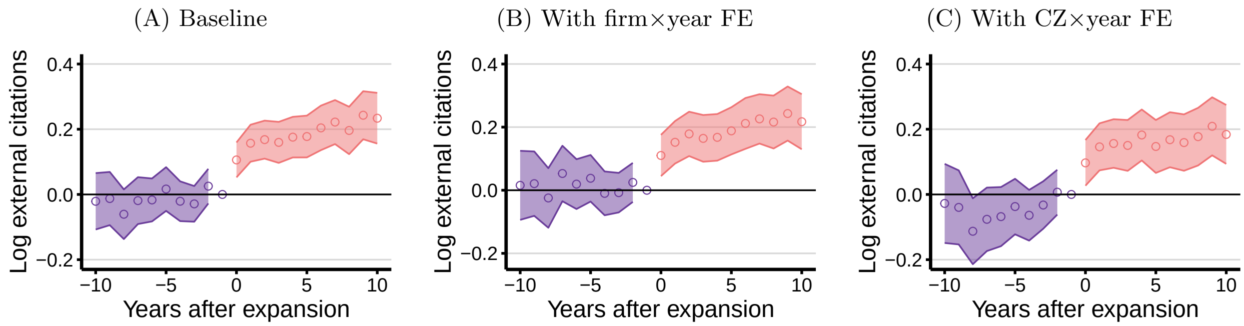
We adopt an event study specification, leveraging the staggered expansion of firms into new markets:

$$y_{\omega nt} = \mathbf{FE}_{\tilde{g}_{\omega n}\omega n} + \mathbf{FE}_{\tilde{g}_{\omega n}t} + \sum_{\tau \in \{-10,\dots,10\}\setminus\{-1\}} \delta_\tau \mathbb{I}\{t - g_{\omega n} = \tau\} + \varepsilon_{\omega nt}, \quad (13)$$

where $y_{\omega nt}$ is the outcome of interest for firm ω in commuting zone n at time t, $\mathbf{FE}_{\tilde{g}_{\omega n}\omega n}$ and $\mathbf{FE}_{\tilde{g}_{\omega n}t}$ are cohort-specific unit and time fixed effects, δ_τ are the coefficients of interest measuring the effect of expansion at horizon τ relative to the year before entry ($\tau = -1$), and $g_{\omega n}$ is the first year in which firm ω is observed in commuting zone n.[19] We use first observed establishment presence rather than first local patenting so that expansion can precede measured patent output in the new market. Comparing expansions across different types of firms would likely violate the parallel trends assumption due to fundamentally different firm trajectories and market conditions. To identify the effect of expansion, we therefore implement a within-firm matching approach. For example, we compare citations of IBM in a market in which it expanded (the treated unit) to a

[19] Control units are firm-market pairs where firm ω is never active in commuting zone n (i.e., $g_{\omega n} = \infty$). Each control unit is assigned the cohort of its matched treated unit, so that $\tilde{g}_{\omega n} = g_{\omega n}$ for treated units and $\tilde{g}_{\omega n}$ equals the matched treated unit's cohort for controls.

Figure 3: **Firm expansion increases local external citations**



Note: The figure shows event study estimates of equation (13), where the dependent variable is log external citations received by the expanding firm from other firms in the expansion commuting zone. Panel 3A includes cohort-unit and cohort-time fixed effects; Panel 3B adds firm×year×cohort controls; Panel 3C adds CZ×year×cohort controls. Error bands are 95% confidence intervals. Standard errors are clustered at the firm–CZ–cohort level. Wald tests of joint pre-treatment coefficient equality to zero: $p = 0.50$, 0.83, 0.46. Point estimates and standard errors are reported in Table D.1.

market in which it did not expand (the control unit). Each expansion market is matched to a non-expansion market based on pre-treatment outcomes – both the level and rate of change – and pre-treatment covariates including commuting zone demographics and firm dynamics.[20] Further details of the matching procedure can be found in Appendix C.

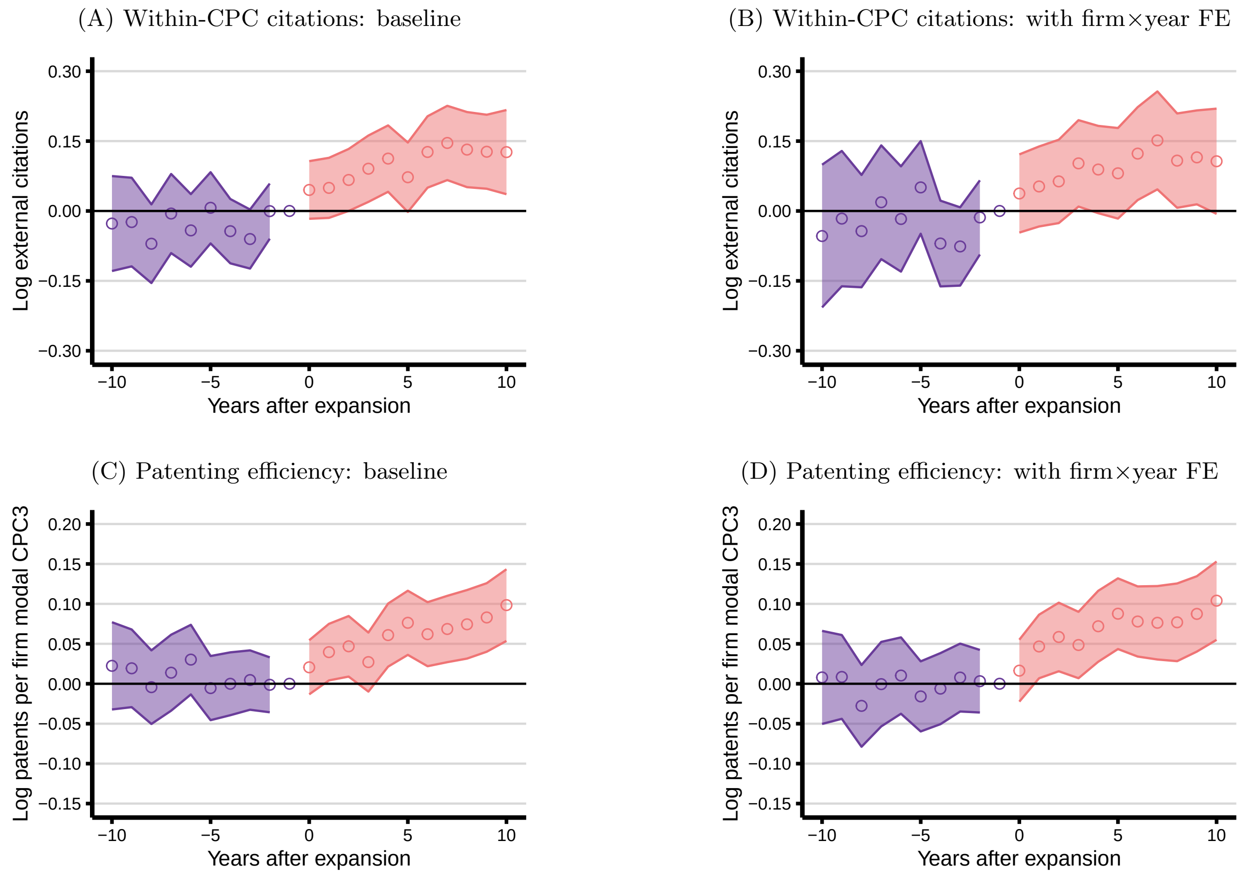
Figure 3 presents our results for external citations. The dependent variable is the log number of external citations received by firm ω's patents from other firms in commuting zone n.[21] Panel 3A shows that firm expansion into a market is followed by a significant and persistent increase in external citations. Ten years following expansion, citations are roughly 25% higher in treated commuting zones compared to control commuting zones, with minimal evidence of pre-trends or anticipation. To address the concern that time-varying firm-level shocks – such as changes in a firm's innovation strategy or patenting propensity – may drive the result, Panel 3B adds firm×year fixed effects; the estimates are essentially unchanged. To address the concern that time-varying local shocks – such as changes in regional R&D policy or local economic conditions – may generate a spurious correlation between firm entry and citations, Panel 3C adds commuting zone×year fixed effects; the result remains robust.

The citation results above are consistent with the presence of knowledge flows between the expanding firm and other local inventors. We now ask whether these patterns are also accompanied by innovation gains for other local firms, as measured by patenting efficiency (patents per firm). Unlike citations, which vary at the firm×market level, patenting efficiency of local firms is defined

[20] This approach is a variant of the "stacking" estimator adopted in Cengiz et al. (2019). As discussed in Roth et al. (2023), Gardner (2022) shows that this framework estimates a convex weighted average of cohort-specific effects under parallel trends and no anticipation.

[21] These are citations received from the patents of other firms $\omega' \neq \omega$ to patents of reference firm ω. We show in the appendix that our results are essentially identical if we add two further restrictions to our citation series: (1) We count only citations to patents of firm ω invented by ω's inventors located outside the expansion commuting zone. (2) We count only citations made by patents of other firms $\omega' \neq \omega$ whose inventors never patent for firm ω (e.g., no shared inventors).

Figure 4: **Firm expansion raises local patenting efficiency**



Note: The figure shows event study estimates of equation (13) for two outcomes. Panels 4A and 4B: log external citations received by the expanding firm from same-CPC patents in the expansion commuting zone. Panels 4C and 4D: patents per firm in the expanding firm's dominant CPC class in market n, excluding the expanding firm. Panels 4A and 4C use baseline fixed effects; Panels 4B and 4D add firm×year×cohort controls. Error bands are 95% confidence intervals. Standard errors are clustered at the firm–CZ–cohort level. Wald tests of joint pre-treatment coefficient equality to zero: $p = 0.38$, 0.37, 0.73, 0.80. Point estimates are reported in Table D.1.

at the market level, which limits the available sample size and variation. We therefore investigate this relationship using a narrower definition of a market – a commuting zone×CPC cell – tracing how patenting efficiency in the expanding firm's dominant technology class changes following its entry into a new market.[22] This allows us to measure patenting efficiency among technologically proximate firms while excluding the expanding firm itself.

Figure 4 shows the results. We first replicate the citation analysis at this finer level: Panels 4A and 4B show that 10 years following expansion, citations from same-CPC patents are approximately 10–13% higher, an effect that is robust to inclusion of firm×year fixed effects. We then turn to patenting efficiency: Panels 4C and 4D show that patents per firm in the local technology cluster rises by around 10%, consistent with the idea that the citation patterns documented above are

[22] Cooperative Patent Classification (CPC) codes are a standardized system used by patent offices to categorize patents by technological content. Patents receive one or more CPC codes; in our analysis we use the primary CPC.

accompanied by economically meaningful gains in the innovative output of nearby firms.[23] As before, results show minimal evidence of pre-trends.

Additional specifications and robustness exercises can be found in Appendix D, including additional controls (Figure D.1); static difference-in-differences specifications (Figure D.2); exclusion of firms with supply-chain relationships, to rule out demand-driven co-location (Figure D.3); more restrictive citation definitions, to address mechanical citation channels (Figure D.4); alternative estimators including PPML, to account for zeros in citation counts (Figure D.5); restriction to persistent expansions, to ensure results are not driven by temporary entry (Figure D.7); and estimation on the million dollar plant sample of Greenstone et al. (2010), which while offering limited variation due to the small sample of patenting expanding firms, provides more plausibly exogenous expansions (Table D.2). Appendix D also explores heterogeneity in spillovers along additional dimensions: Figure D.8 splits expansions by whether the entrant patents locally within five years, finding larger spillovers when the new location becomes an active patenting site; and Figure D.9 partitions by whether the firm's technology matches the local market's dominant CPC, finding similar effects across the two groups – consistent with the evidence in Figure 2B that firms replicate their innovation portfolios across markets rather than tailoring them to local specialization.

3.4 Spillovers and local firm size

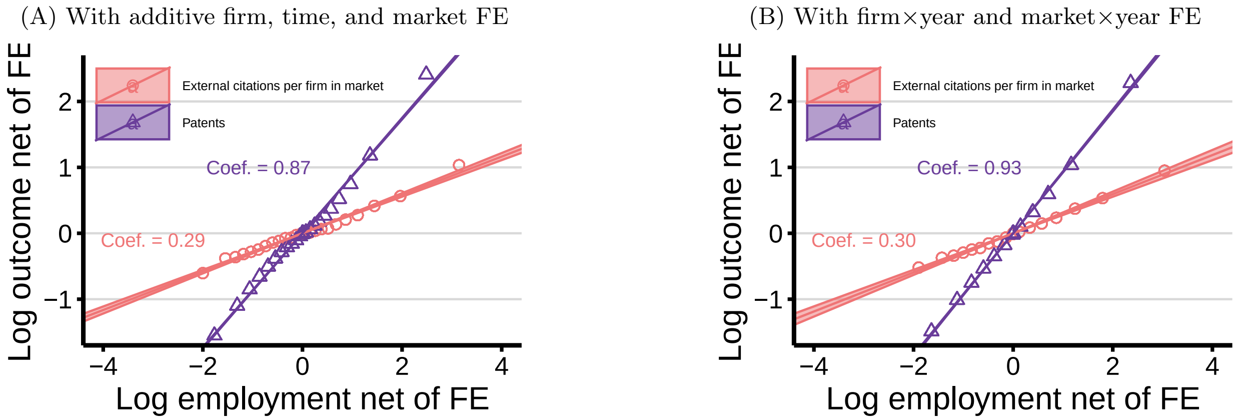
The evidence presented so far suggests that firm expansion generates local knowledge spillovers, as measured by patent citations. We now examine how these spillovers scale with a firm's local employment. We also compare this to how the firm's own innovation output scales with local employment. This comparison is central to the model's prediction about whether firms operate in too few or too many markets: if spillovers are less sensitive to local firm size than innovation output is ($\beta < \eta$), the model implies that firms under-expand.

To examine this, we compare how two distinct outcomes produced by firm ω in market n co-vary with the firm's local innovation employment. The first is external patent citations received by firm ω from other firms in market n, which proxies for the spillovers that firm ω generates locally. The second is the number of patents filed by firm ω in market n, which proxies for the firm's own local innovation output.

Figure 5 presents the results. In both panels, the slope of the patents-employment relationship (blue) is steeper than that of the citations-employment relationship (red). In Panel 5A, which controls for additive firm, time, and market fixed effects, a firm's local patenting output rises with an elasticity of 0.93 with respect to its local employment, while external citations rise with a substantially lower elasticity of 0.30. Panel 5B shows that this pattern is robust to controlling for firm×year and market×year fixed effects, which absorb both time-varying firm-level productivity shocks and time-varying local conditions. In Appendix E, we report highly similar results using

[23] For these exercises, we do not include a cohort-commuting zone-year fixed effect for robustness. Because the outcome is constructed at the commuting zone-technology class-year level, with only the focal firm left out, it is nearly-mechanically absorbed by such fixed effects. The leave-out procedure generates residual firm-level variation, but this is too thin to survive the inclusion of a saturated cohort-commuting zone-year control.

Figure 5: **Local innovation output and spillovers scale differently with firm size**



Note: Binscatter plots of log local patents (blue triangles) and log local external citations received (red circles) against log local inventor employment, for the same firms in the same markets. Dots are 25 equal-sized bins, net of fixed effects. Panel 5A: additive firm, time, and commuting zone fixed effects. Panel 5B: firm×year and market×year fixed effects. Error bands are 99% confidence intervals. Standard errors are clustered at the firm×market level.

different measures of a firm's local innovation labor, including lagged inventor counts (Figure E.3), which address the mechanical correlation between contemporaneous inventors and patenting output, and Dun & Bradstreet employment (Figure E.4).

The key pattern is that a firm's local innovation output is more responsive to its local employment than the spillovers it generates for other firms. To interpret these slopes through the lens of the model, recall that regional R&D productivity is given by $K_{nt} = \bar{K}_n \left(\int_{\omega \in \Omega_{nt}} \iota_{\omega nt} \, d\omega \right)^{\alpha}$, where $\iota_{\omega nt} \equiv z_\omega \ell_{\omega nt}^{\beta}$ denotes firm ω's contribution to local knowledge in market n at time t. Intuitively, $\iota_{\omega nt}$ captures the rate at which ideas generated by firm ω arrive to other firms operating in market n. We proxy $\iota_{\omega nt}$ using the average number of citations that firm ω receives from firms in market n at time t, so that the elasticity of external citations with respect to local employment (the slope of the citations regression line in Figure 5) informs β. Similarly, the firm's own local innovation output is $\lambda_{\omega nt} = \lambda_0 K_{nt} z_\omega \ell_{\omega nt}^{\eta}$, so that the elasticity of patenting with respect to local employment (the slope of the patents regression line) informs η. The fact that the patenting slope is consistently steeper than the citations slope thus provides a first indication that $\beta < \eta$: the marginal worker in a given firm contributes more to that firm's own innovation than to the knowledge available to other local firms – consistent with strong diminishing returns in the spillovers that a firm generates locally from the expansion of its local activity. Through the lens of the model, this implies that firms under-expand in equilibrium. In Section 4, we return to these relationships and estimate β and η more formally, addressing endogeneity and measurement concerns.

4 Model estimation

We now turn to developing a quantitative version of the model to address the following questions. First, do innovative firms operate in too few or too many markets, that is, is β low enough relative to η? Second, what are the potential welfare gains from policies that influence firms' spatial expansion, particularly in comparison to more conventional interventions such as R&D subsidies? Third, given the spatial heterogeneity in the incidence of such policies, what are their implications for regional inequality? We first extend the baseline model by introducing additional components that make it suitable for quantitative analysis. We then estimate the model to the U.S. economy in 2018, mapping regions in the model to U.S. labor market areas (LMAs, aggregations of commuting zones) and aggregating small LMAs to a single region, resulting in 300 local markets.

4.1 Quantitative model

We enrich the model with additional components that help to discipline the baseline equilibrium, and which are standard in the economic geography literature.

Labor Mobility. Households choose both where to live and which sector to work in either innovation or production. To preserve tractability, we assume logarithmic flow utility, $\phi = 1$. We focus on a balanced growth path equilibrium, and assume that agents choose a location and sector once in an ex-ante stage, taking into account that the economy is on a balanced growth path. Choices are subject to additive idiosyncratic preference shocks for locations and sectors, drawn from a Type-1 Extreme Value distribution. Under these assumptions, agents choose location and sector to maximize

$$\max_{n\in\{1,\ldots,N\},\, s\in\{p,i\}} \frac{1}{\rho}\log C_{ns} + \frac{g}{\rho^2} + \varepsilon_{ns}, \quad \varepsilon_{ns} \sim \text{Gumbel}\left(\frac{1}{\rho\epsilon}, \log B_{ns}^{\frac{1}{\rho\epsilon}}\right),$$

where $\frac{1}{\rho\epsilon}$ and $\log B_{ns}^{\frac{1}{\rho\epsilon}}$ are the scale and location parameters governing the distribution of preference shocks for sector s in location n, and C_{ns} is the detrended relative consumption if choosing sector s and location n. The parameters B_{ns} capture location-sector amenities or other forms of compensating differentials.

Local housing. We now assume that agents consume both a national traded good and a local good in the form of housing, introducing an additional congestion force and enabling realistic variation in local prices. The consumption bundle is a Cobb-Douglas aggregator of tradable goods and housing, with expenditure shares $1-\varpi$ and ϖ, respectively. Each location is endowed with one unit of housing, and the local rental rate is denoted by R_{nt}. We assume that all housing rents, like monopolistic profits, accrue to a national fund and are redistributed to households in proportion to their labor income. Letting ς denote the aggregate ratio of total income to the aggregate wage bill, households' detrended relative consumption is given by

$$C_{ns} = \varsigma R_n^{-\varpi} w_{ns},$$

where R_n and w_{ns} are the detrended rental rate and wage, respectively.

Under the above assumptions, the share of agents choosing location n and sector s takes the standard choice probability form:

$$\frac{B_{ns}\left(R_n^{-\varpi} w_{ns}\right)^{\epsilon}}{\sum_{n'=1}^{N} \sum_{s' \in \{i,p\}} B_{n's'}\left(R_{n'}^{-\varpi} w_{n's'}\right)^{\epsilon}}.$$

With this formulation, ex-ante welfare – i.e., expected utility before idiosyncratic preference shocks are realized – is given by

$$U = \frac{1}{\rho} \log \left(\sum_{n'=1}^{N} \sum_{s' \in \{i,p\}} B_{n's'}\left(R_{n'}^{-\varpi} \varsigma w_{n's'}\right)^{\epsilon} \right)^{\frac{1}{\epsilon}} + \frac{g}{\rho^2}. \tag{14}$$

4.2 Estimation of key parameters

We now describe how we discipline the key parameters that govern the novel aspects of the model: α, β, η, and θ. We employ a moment matching approach, using four moments from the data that jointly identify these four parameters. While the quantitative model is solved at the LMA level, the moments that identify α, β, η, and θ are estimated from finer annual variation in CZ-year and firm-by-CZ-year observations.

4.2.1 Regional spillovers

Similar to the literature on regional agglomeration, our model implies a relationship between regional scale (as measured by total innovation labor, H_n) and the innovation productivity of local firms (Greenstone et al., 2010; Giroud et al., 2026; Moretti, 2021). We use this key relationship as one of our moments to identify our main parameters. Proposition 3 in Section 2 highlights that an increase in H_n, holding constant the exogenous component of regional innovation productivity, $\bar{K}_n$, leads to an increase in firm-level innovation output, with the elasticity given by:[24]

$$\frac{\partial \log\left(\lambda_n / M_n\right)}{\partial \log\left(H_n\right)} = \alpha \left[\frac{\theta}{1+\theta} + \frac{\beta}{1+\theta}\right] + \frac{\eta}{1+\theta}, \tag{15}$$

The first term captures the benefit of agglomeration due to regional spillovers when $\alpha > 0$: β governs the strength of spillovers operating through the increase in the size of local firms, while $\frac{\theta}{1+\theta}$ reflects spillovers arising from an increase in the mass of local firms. The second term reflects the direct effect of higher innovation labor within the firm, governed by the decreasing returns to scale parameter, η. While in the agglomeration literature this relationship is typically captured by a single parameter that determines the strength of agglomeration forces (α), our model introduces additional channels. Lower β and η dampen the effect of regional scale on innovation productivity due to diminishing returns in the spillover function and in the production of ideas, operating as

[24] This expression is derived in appendix A.4.

counter-forces to α. Higher θ can amplify or dampen the scale elasticity, as it implies on the one hand more firms to learn from, and on the other hand more resources soaked into payment of fixed operating costs relative to generation of ideas. In the limiting case of $\theta \to \infty$, in which all firms are active in all markets and there is no tradeoff between opening new plants and expanding existing ones, the scale elasticity equals α as in the benchmark agglomeration model.

To identify the scale elasticity, we estimate the following specification:

$$\log\left(\frac{\lambda_{nt}}{M_{nt}}\right) = \mathbf{FE}_n + \mathbf{FE}_t + \tilde{\alpha}\log H_{nt} + \varepsilon_{nt}, \tag{16}$$

where λ_{nt} is the number of patents in region n at time t, M_{nt} is the number of innovative firms in region n at time t, H_{nt} denotes the local innovation labor force, $\mathbf{FE}_n$ are region fixed effects, $\mathbf{FE}_t$ are year fixed effects, and ε_{nt} is a residual.[25] Note that when one uses variation in H_{nt} that is not due to shocks to $\bar{K}_n$ (e.g., shocks to regional labor supply), the estimated value of $\tilde{\alpha}$ corresponds to the composite of model parameters $\alpha\left[\frac{\theta}{1+\theta} + \frac{\beta}{1+\theta}\right] + \frac{\eta}{1+\theta}$.

We can also investigate the same relationship at the firm level, using within-firm variation, by estimating the following specification:

$$\log\left(\lambda_{\omega nt}\right) = \mathbf{FE}_n + \mathbf{FE}_{\omega t} + \tilde{\alpha}\log H_{nt} + \varepsilon_{\omega nt} \tag{17}$$

where $\lambda_{\omega nt}$ is now firm-level patenting in region n, and $\mathbf{FE}_{\omega t}$ is a firm$\times$year fixed effect. In this case, $\tilde{\alpha}$ is the partial elasticity of firm-level innovation output with respect to regional innovation labor, holding constant regional productivity ($\bar{K}_n$) and firm productivity (z_ω). Note that both empirical specifications recover the same structural parameters, which measure the responsiveness of average regional innovation to changes in regional innovative employment.

A natural challenge for identification is the presence of confounding, unobserved regional factors in the error term, captured in the model by shocks to $\bar{K}_{nt}$, which would lead to both higher local innovation activity and to changes in H_{nt}. A second natural concern is the measurement error in H_{nt}, since one needs to take a stance on the correct measure of local innovation labor.

To address these concerns, we use an instrumental variables strategy that shifts the supply of innovative labor in a location without directly affecting innovation efficiency. To this end, we adopt the predicted tax instrument of Akcigit et al. (2022). This instrument exploits top federal personal income tax changes and interacts them with predetermined state tax schedules and deductibility rules, holding state tax parameters fixed at their lagged levels. Identification comes from federally-driven (i.e., non-local) variation in the personal income tax rates facing inventors. Through the lens of our model, this acts as a shift in the supply of innovative labor (i.e. a change in the labor supply shifters B_{nit}), holding constant the exogenous component of regional innovation productivity, $\bar{K}_n$.

Another possible concern is that counts of patents and inventors are somewhat mechanically linked, as we observe inventors only when they file a patent. To ensure this does not drive our results, we adopt a split-sample approach to estimating (16). We randomly partition the firms into

[25] Because the model is estimated along a (BGP), all regressions use time fixed effects to absorb time variation.

Table 2: **Estimation of the regional scale elasticity.**

	Panel A: Region-level			Panel B: Firm-level		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	IV (split)	OLS	IV	IV (leave-out)
$\log H_{nt}$	0.75	0.51	0.44	0.57	0.66	0.65
	(0.03)	(0.16)	(0.21)	(0.04)	(0.19)	(0.20)
N	7420	7420		154518	154518	154015
KP F-stat.		18.17	12.65		54.12	53.10
CZ FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓			
Firm×Year FE				✓	✓	✓

Note: Panel A reports estimates of region-level regression (16); Panel B reports estimates of firm-level regression (17). The reported coefficient, $\tilde{\alpha}$, is the elasticity of region-level patents per firm in Panel A and of firm-level patenting in Panel B with respect to local innovation employment, holding fixed $\bar{K}_n$. Panel A is estimated on commuting zone-year cells with more than four active firms, so that the split-sample construction is well defined; we apply the same restriction across columns (1)–(3) to keep the underlying regional sample comparable across specifications. Column (3) uses the split-sample IV procedure described in footnote 26, and column (6) uses leave-out IV, which excludes unit ωnt's contribution to H_{nt}. In Panel A, parentheses report standard errors clustered at the commuting zone level. In Panel B, parentheses report standard errors two-way clustered at the firm and year levels.

two groups: one used to calculate patents per firm and the other used to calculate market-level inventor employment.[26]

The estimation results are reported in Table 2. Columns (1)–(3) present the regional-level estimates using OLS, IV, and split-sample IV, respectively, while columns (4)–(6) report the firm-level estimates using OLS, IV, and leave-out IV. For leave-out IV, we simply exclude unit ωnt's contribution to H_{nt} when estimating the IV regression.

We obtain estimated elasticities in the range 0.44-0.66 using our instrumental variables strategy.[27] All coefficients are statistically significant at the 5% level, and the first-stage F-statistics are above the conventional rule-of-thumb threshold of 10. For our quantitative analysis, we choose the most conservative estimate for the scale elasticity, delivered by the split-sample regional IV with a coefficient of 0.44. This is our first estimating moment, which contains information about α, but also about β, η, and θ. In the next sections, we discuss how to separate these different parameters.

[26] As discussed in Appendix B.2, when allocating inventors to firms, we impute inventor presence in years without patenting activity, but can only do so for inventors who patent in multiple years. This creates an unavoidable mechanical correlation between inventor counts and patent counts. To address this concern, we adopt a split-sample approach; we repeat this procedure 50 times and report the average coefficient and standard error across bootstrap iterations. We show a density plot of these regression coefficients in Figure E.1.

[27] Note that this is the elasticity of average regional innovation output *per firm* with respect to regional innovation labor. To obtain the elasticity of *total innovation output* with respect to regional innovation labor, one would need to add to it the elasticity of the mass of regional innovating firms with respect to regional innovation labor, which is given in the model by $\frac{\theta}{1+\theta}$. In spatial models featuring constant returns to scale, this second elasticity is always 1 by construction.

Table 3: **The responsiveness of local firm size to regional innovation labor supply.**

	(1)	(2)	(3)
	OLS	IV	IV (leave-out)
$\log H_{nt}$	0.44	0.50	0.49
	(0.03)	(0.16)	(0.16)
N	278775	278775	277511
KP F-stat.		56.41	57.46
Firm×Year FE	✓	✓	✓
CZ FE	✓	✓	✓

Note: The table reports estimates of Equation 18. The reported coefficient, $\tilde{\theta} = 1/(1+\theta)$, is the elasticity of firm-level local employment with respect to total market-level employment. Column (3) uses leave-out IV, which excludes unit ωnt's contribution to H_{nt}. Parentheses report standard errors two-way clustered at the firm and year levels.

4.2.2 Estimation of θ

The second moment that we use allows us to identify θ, the dispersion in firms' fixed costs across locations. As we show in Appendix A.4, we can exploit the same variation and instrumental variables strategy as in Section 4.2.1, but now study the response of average firm-level local employment in market n to shocks to total regional employment H_{nt}, which is proportional in the model to

$$\ell_{\omega n} \propto z_{\omega}^{\frac{1}{1-\eta}} H_n^{\frac{1}{1+\theta}}.$$

An increase in total regional innovation labor can translate either to more local plants or to larger local plants. Due to variation in fixed costs, marginal plants require a higher fixed cost relative to the average plant in a location, which shapes the degree to which an increase in regional employment translates into an increase in firm local employment. Therefore, tracing the response of average firm size (or equivalently, the local mass of firms) to changes in local innovation labor supply allows us to identify θ.

To this end, we estimate the following regression:

$$\log \ell_{\omega nt} = \mathbf{FE}_n + \mathbf{FE}_{\omega t} + \tilde{\theta} \log H_{nt} + \varepsilon_{nt} \tag{18}$$

where $\mathbf{FE}_n$ is a region fixed effect, and $\mathbf{FE}_{\omega t}$ is a firm-year fixed effect which absorbs firm-level productivity, z_ω. Through the lens of our model, a change in H_{nt} facilitated by labor supply shocks identifies $\tilde{\theta} = 1/(1+\theta)$. While in principle an OLS regression with firm×year fixed effects would be sufficient to identify $\tilde{\theta}$ through the lens of our model, we also estimate it using the same IV strategy as in Section 4.2.1. This allows us to deal with similar concerns regarding measurement error in H_{nt}, and to address concerns that are not explicitly modeled, such as a correlation between unobserved shocks to $\bar{K}_n$ and changes in the fixed costs of operating plants in that location.

Table 3 reports our results. All estimates are significant at the 1% level and first-stage F-statistics are above the conventional rule-of-thumb threshold of 10. Estimates are largely stable across specifications. For the quantitative analysis, our preferred specification is the leave-out IV regression, yielding $\theta = 1.04$.

4.2.3 Estimation of η and β

We now describe how two additional moments allow us to jointly pin down the parameters η and β, which capture diminishing returns to scale in the production of ideas and the local spillovers that firms generate for other firms, respectively. We have already seen in Section 3.4 evidence that a firm's local innovation output is more responsive to its local employment than the spillovers it generates for other firms (Figure 5), suggesting that $\beta < \eta$. We now estimate both parameters using model-consistent specifications.

Estimating the difference between η and β. Recall the functional forms for the ideas produced by a firm in a given market and for the spillovers it generates for other firms in that market:

$$\log(\iota_{\omega nt}) = \log(z_\omega) + \beta \log(\ell_{\omega nt}) \tag{19}$$

$$\log(\lambda_{\omega nt}) = \log(\lambda_0) + \log(K_{nt}) + \log(z_\omega) + \eta \log(\ell_{n\omega t}), \tag{20}$$

where $\lambda_{\omega nt}$ is the arrival rate of ideas for firm ω in market n, and $\iota_{\omega nt}$ is the arrival rate of insights from that firm to other firms in market n. Taking the difference between them, we obtain:

$$\log(\lambda_{\omega nt}/\iota_{\omega nt}) = \log(\lambda_0) + \log(K_{nt}) + (\eta - \beta)\log(\ell_{\omega nt}). \tag{21}$$

Equation 21 suggests that by holding fixed the regional labor productivity K_{nt}, we can estimate $\eta - \beta$ from the responsiveness of a firm's local output-to-spillovers ratio to its local size. Importantly, identification requires variation in a firm's local size that stems from firm-level shocks (e.g., a change in z_ω), not from regional changes in K_{nt}, for example by comparing firms of different sizes in the same local market. If a firm's innovation output responds more than its spillovers to others when the firm increases its local size, as suggested by the evidence in Section 3.4, then $\eta - \beta > 0$.

We operationalize this by estimating the following equation:

$$\log(\lambda_{\omega nt}/\iota_{\omega nt}) = \mathbf{FE}_{nt} + (\eta - \beta)\log(\ell_{\omega nt}) + \varepsilon_{\omega nt}, \tag{22}$$

where $\mathbf{FE}_{nt}$ denotes region×time fixed effects that absorb K_{nt}. We proxy $\iota_{\omega nt}$ with average realized external citations from firms in region n and $\lambda_{\omega nt}$ with local patents. Several identification concerns remain. First, our measure of a firm's local innovation labor, $\ell_{\omega nt}$, is derived from patent data, which captures only inventors who appear on patent filings and may therefore miss part of a firm's total local R&D workforce. Relatedly, because both $\lambda_{\omega nt}$ and $\ell_{\omega nt}$ are deduced from patent data, there is a potential mechanical correlation between the dependent variable and the regressor.

Table 4: **The elasticity of patents-to-citations ratio to local size.**

	Panel A: Employment			Panel B: Lagged employment		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	IV	OLS	OLS	IV
$\log \ell_{\omega nt}$	0.43	0.84	0.64	0.33	0.42	0.56
	(0.02)	(0.01)	(0.09)	(0.01)	(0.02)	(0.14)
N	147011	147011	147011	147011	147011	147011
KP F-stat.			17.01			15.42
p-Value $H_0 : \eta - \beta \leq 0$	0.00	0.00	0.00	0.00	0.00	0.00
CZ×Year FE	✓	✓	✓	✓	✓	✓
Firm×CZ FE		✓	✓		✓	✓

Note: The table shows results from a regression of log patents minus log average citations on firm-level local employment (Equation 22). We use two definitions of firm-level local employment: inventor counts (Panel A) and lagged inventor counts (Panel B). Parentheses report standard errors clustered at the firm level.

Second, while the model does not allow a firm to be idiosyncratically more productive in some regions than in others, this may be a feature of the data. Third, unmodeled shocks may jointly shift a firm's local employment and the citations it receives without stemming from a change in the firm's productivity. For instance, a wave of local enthusiasm about a firm – heightened excitement about its prospects – could increase labor supply to ω relative to other firms in market n while simultaneously raising citations to that firm.

To address these concerns, we employ the following steps. First, we augment the specification with firm×CZ fixed effects to absorb persistent idiosyncratic matches between a firm and a region. Note that we do not add firm×year fixed effects: through the lens of the model, firm-level productivity shocks are the main source of variation that shifts ω's local employment relative to other firms in the same market, and absorbing them would eliminate the identifying variation. Second, we consider both contemporaneous and lagged inventor counts as our measures of $\ell_{\omega nt}$, to mitigate the mechanical correlation discussed above. Third, to address remaining measurement error and unmodeled shocks such as the local-enthusiasm channel, we instrument $\ell_{\omega nt}$ with firm ω's log inventor counts in all other markets, excluding n. This leave-one-out instrument isolates the component of local employment driven by firm-level productivity while purging the influence of location-specific shocks.

Table 4 reports the results. In Panel A, we use contemporaneous inventor counts as our measure of $\ell_{\omega nt}$. Column (1) presents the baseline OLS estimate: $\eta - \beta = 0.43$. Column (2) adds firm×CZ fixed effects to absorb persistent firm-region heterogeneity, yielding a larger estimate of 0.84. Column (3) instruments $\ell_{\omega nt}$ with the leave-one-out measure while including both region×year and firm×CZ fixed effects: $\eta - \beta = 0.64$, statistically significant at the 1% level, with a first-stage F-statistic of 17.01. Panel B replicates all three specifications using lagged inventor counts to mitigate the mechanical correlation between contemporaneous employment and patenting. The resulting estimates – 0.33, 0.42, and 0.56 – are somewhat smaller but remain highly statistically significant

Across all specifications, one-sided t-tests consistently reject $\eta - \beta \leq 0$ at greater than 0.1% significance, confirming that patenting co-moves with local employment more strongly than citations, as suggested by Figure 5. Our preferred estimate for the quantitative analysis is the IV specification with lagged employment (column 6), delivering $\eta - \beta = 0.56$.

Estimation of β. Finally, we need to determine either η or β individually to pin down all key parameters. We estimate β directly from equation (19), exploiting within-firm variation in local employment across markets. Comparing the same firm in the same year across its different locations, we study how the external citations it receives in each market depend on its local size. We estimate the following equation:

$$\log(\iota_{\omega nt}) = \mathbf{FE}_{\omega t} + \beta \log(\ell_{\omega nt}) + \varepsilon_{\omega nt}, \tag{23}$$

where $\mathbf{FE}_{\omega t}$ denotes firm×year fixed effects that absorb firm-level productivity z_{ω}, and we continue to proxy $\iota_{\omega nt}$ with average realized external citations from firms in region n. Through the lens of the model, variation in $\ell_{\omega nt}$ across markets for a given firm stems from differences in local market conditions, and the firm×year fixed effects absorb firm productivity.

Several identification concerns are worth discussing, mirroring those from the previous section. First, unlike regressions that include patents as part of the outcome variable, here external citations are the outcome. This eliminates the mechanical correlation between local employment and the dependent variable, since ω's inventor counts are derived from its patents while external citations come from other firms' patents. Second, measurement error in $\ell_{\omega nt}$ remains a concern, since our inventor-based employment measure may miss part of a firm's local R&D workforce. To address this, we can instrument $\ell_{\omega nt}$ with an alternative measure of local employment from an independent data source, such as D&B employment. As long as the measurement errors across these sources are uncorrelated – plausible given their completely different origins – this eliminates attenuation bias from mismeasurement of $\ell_{\omega nt}$.

Third, while the above specification is valid under the model's assumptions, one might worry that regional shocks facilitating the change in the firm's employment in market n (e.g., a shock to $\bar{K}_{nt}$) could independently generate a surge in citations to that firm. This was less of a concern when estimating $\eta - \beta$, where region×year fixed effects absorbed all regional variation and identification relied on within-market cross-firm comparisons. Here, identification relies on within-firm variation across markets, so we cannot include region×year fixed effects without absorbing the identifying variation.

To address this concern, we provide additional evidence using our event-study design from Section 3. By comparing expansions of different sizes, we can trace the differential response of citations to larger versus smaller expansions. Specifically, we repeat specification (13), splitting the sample between the largest 50% and smallest 50% of expansions, where expansion size is measured

Table 5: **Regression of external citations per firm on local labor.**

	(1)	(2)	(3)	(4)	(5)	(6)
$\log \ell_{\omega nt}$	0.30	0.33	0.35	0.21	0.15	0.15
	(0.01)	(0.03)	(0.03)	(0.09)	(0.11)	(0.24)
N	196055	196055	196055			
KP F-stat.		18.87	17.35			
Firm×Year FE	✓	✓	✓		✓	
CZ×Year FE			✓			✓

Note: Columns (1)–(3) report cross-sectional estimates of Equation (23). Column (1) is the model-consistent OLS specification with firm×year fixed effects. Column (2) reports the corresponding TSLS estimate, which addresses measurement error in inventor-based employment while preserving the same identifying variation. Column (3) adds CZ×year fixed effects to address unmodeled location-year shocks. The reported Kleibergen-Paap F-statistics refer to the corresponding first stages. Columns (4)–(6) report event-study estimates of the approximation in Equation (24), mirroring the progression of event-study specifications in Figure 3. Column (4) is the baseline event-study specification, column (5) adds cohort×firm×year fixed effects, and column (6) adds cohort×CZ×year fixed effects. Parentheses report standard errors clustered at the firm level in columns (1)–(3) and bootstrap standard deviations in columns (4)–(6). Figure E.2 plots the bootstrap distributions underlying columns (4)–(6).

as the average number of the firm's inventors in the new market during the first four years of presence. As in Section 3, we present three specifications: a baseline with cohort-specific unit and time fixed effects, an augmented specification with firm×year fixed effects to absorb time-varying firm-level shocks, and a specification with CZ×year fixed effects to absorb time-varying local shocks. We then approximate β using the relation:

$$\beta \approx \frac{\bar{\delta}_\tau^{\text{Large}} - \bar{\delta}_\tau^{\text{Small}}}{\overline{\log \ell}^{\text{Large}} - \overline{\log \ell}^{\text{Small}}}, \tag{24}$$

where $\bar{\delta}_\tau^{\text{Large}}$ and $\bar{\delta}_\tau^{\text{Small}}$ denote the average post-treatment difference-in-differences coefficients for large and small expansions, and the denominator is the difference in mean log expansion size across the two groups. This approximation recovers the elasticity of external citations with respect to local labor by comparing how much more citations increase following larger versus smaller employment expansions, scaled by the corresponding difference in log local labor.

Table 5 reveals three related patterns. First, the cross-sectional estimates in columns (1)–(3) are stable across specifications. Column (1) is the model-consistent specification: conditional on firm×year fixed effects, within-firm differences in local employment across markets identify how spillovers scale with local footprint. It yields $\beta = 0.30$. Column (2) keeps that same comparison but instruments inventor-based employment with D&B employment to address measurement error, and the estimate rises only modestly to 0.33. Column (3) uses the same D&B-employment instrument and adds CZ×year fixed effects to absorb unmodeled location-year shocks, and the estimate remains positive at 0.35. Thus, the positive relationship between local scale and external citations is not fragile to either measurement-error corrections or to absorbing common regional shocks. At the same time, even the largest cross-sectional estimate remains far below the unit elasticity that would arise if spillovers scaled one-for-one with local firm size.

Second, columns (4)–(6) provide a complementary dynamic check based on discrete expansion episodes. The baseline event-study estimate in column (4) is 0.21. Adding cohort×firm×year or cohort×CZ×year fixed effects in columns (5) and (6) lowers the point estimate to 0.15 and widens the confidence intervals. These more saturated specifications absorb a large share of the variation around expansion episodes, but the qualitative message is unchanged: all three event-study estimates remain positive and well below one, so the dynamic evidence likewise points to spillover elasticities that are modest relative to the scaling of the firm's own activity.

Third, for the policy counterfactuals the key empirical restriction is that spillovers respond less strongly to local scale than firms' own innovation does, that is, $\beta < \eta$. Our preferred estimate of $\eta - \beta$ from Table 4 is 0.56, and we reject $\beta \geq \eta$ at the 0.1% level. Combining that moment with the baseline event-study estimate $\beta = 0.21$ implies $\eta = 0.77$. We use $\beta = 0.21$ as our preferred direct estimate because it comes from the expansion-based design and is the largest of the three event-study coefficients. In that sense it is a conservative choice for the counterfactuals: using the lower estimates in columns (5) or (6), or even a value closer to zero, the model would move further toward under-expansion. This calibration also lines up closely with an alternative strategy that takes η from the existing literature. Estimates there typically lie between 0.5 and 0.9 (see, e.g., Porter and Stern, 2000; Pakes and Griliches, 1980; Burchardi et al., 2026); using $\eta \approx 0.8$ as in Burchardi et al. (2026) together with our estimate of $\eta - \beta$ yields $\beta \approx 0.24$, very close to our preferred value of 0.21. Thus, whether one relies on our preferred expansion-based estimate or combines our moments with outside evidence on η, the quantitative implications land in essentially the same place.

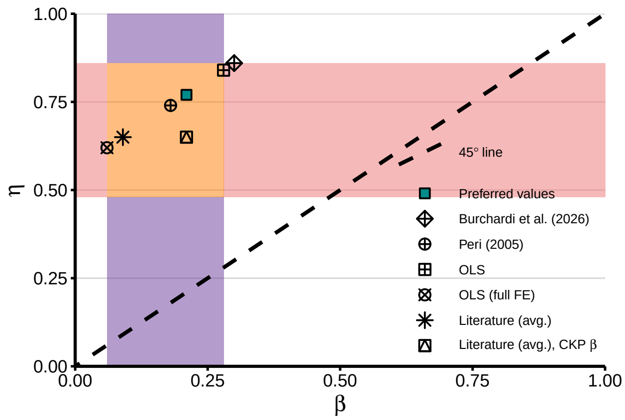
Comparing diminishing returns in spillovers (β) and in idea production (η). Figure 6 summarizes the different combinations of η (plotted on the y-axis) and β (x-axis) that emerge from our estimation, or from combining our independent estimates of β with conventional values of η from the literature. While the approaches differ, a key takeaway is that there is relatively little variation in the results.

4.3 Remaining parameters

Externally Set Parameters. We calibrate the discount rate, ρ, to 0.05. The housing expenditure share is set to 0.25, consistent with evidence from Davis and Ortalo-Magné (2011). The dispersion of workers' preference shocks, ϵ, is fixed at 2.0, aligning with the estimated range for labor supply elasticity across locations and occupations, e.g. in Galle et al. (2023). Similarly, the elasticity of substitution across locations in the production sector, σ, is set to 5.0, following standard values in trade and economic geography literature, such as those in Costinot and Rodríguez-Clare (2014).

Internally calibrated parameters. The remaining parameters we need to determine are the productivity step-size following new innovations, γ; the economy-wide efficiency of innovation, λ_0; the scale of the fixed costs distribution, f_{max}; the cost of entry, f_e; and the efficiency advantage

Figure 6: **Parameter uncertainty for β and η**



Note: The shaded vertical band reports the range of estimated β values, while the shaded horizontal band reports the range of η values considered in the comparison; their overlap is shown in orange. The markers report the alternative parameter pairs listed in the legend. The Peri (2005) marker is based on the 0.74 coefficient on own R&D stock in that paper's innovation regression, so it should be interpreted as a stock-based proxy for η rather than as a direct elasticity with respect to R&D labor. For literature-based comparisons, we use the $\eta - \beta$ estimates from Table 4 to calculate β, except in the case of the asterisk legend item, where we use $\beta = 0.21$ (our preferred β from the event study bootstrap, Table 5).

of large firms, $\bar{z}$. To determine these parameters, we match the following moments: the share of innovation labor in total wage bill (13% in the data); the aggregate growth rate (2%); the average number of markets per innovating firm (2.09); the share of the top 10% of innovative firms in total patenting (70%); and the average employment in patenting firms (70). While these parameters are jointly estimated, we develop an efficient iterative algorithm to compute these parameters using a set of closed-form equations that link these 5 parameters and 5 moments, exactly matching all targeted moments. We provide further information on the computation of these moments and the estimation algorithm in Appendix E.

Inversion of location fundamentals. We invert the regional fundamentals – local productivity in the production sector (Z_n); local productivity in the innovation sector ($\bar{K}_n$); local compensating differentials for innovation workers (B_{ni}); and local compensating differentials for production workers (B_{np}) – by matching data on employment and wages in the production and innovation sectors in all local markets. We provide additional details on the model inversion procedure and how to identify each model residual from data on local employment and wages in Appendix E.2.1.

Table 6 summarizes the parameter values that are used for the quantified model.

5 Gains from incentivizing spatial expansion

We use our estimated model to assess the potential welfare gains of incentivizing innovative firms to expand into more markets. Specifically, we analyze different variants of a spatial expansion subsidy and compare them to a traditional R&D subsidy. Similar to much of the endogenous growth literature, our calibration also leads to under-investment in innovation in equilibrium. Since spatial

Table 6: **Parameter values in the quantified model**

Parameter	Description	Value	Notes
		Externally set	
ρ	Discount rate	0.05	
σ	EoS across locations in production	5	Costinot and Rodríguez-Clare (2014)
ϖ	Housing share of consumption	0.25	Davis and Ortalo-Magné (2011)
ϵ	Dispersion of preference shocks	2	Galle et al. (2023)
		Direct estimation	
η	Elasticity of innovation to R&D inputs	0.77	See text
α	Elasticity of innovation to local spillovers	0.10	
β	Sensitivity of spillovers to local firm activity	0.21	
θ	Dispersion of fixed costs	1.04	
		Internal calibration	
γ	Step size for innovations	1.21	Target: share of innovation in total wage bill
λ_0	Economy-wide innovation efficiency	6.73	Target: growth rate
f_{max}	Scale of fixed costs	371	Target: average # of regions per firm
$\bar{z}$	Efficiency advantage of large firms	1.5	Target: share of top 10% firms in total patents
f_e	Cost of entry into innovation	60.9	Target: average size of patenting firms

expansion subsidies also stimulate innovation activity, they are expected to generate positive welfare gains, even in the absence of local spillovers across innovative firms. By benchmarking these policies against the traditional R&D subsidy, we can evaluate their additional benefits in the presence of local spillovers. We first present the aggregate welfare gains and decompose them into real income and distributional components. We then examine how the ranking of policies varies with the scale of the intervention, and conclude with the regional implications of each policy.

5.1 Aggregate gains

We begin with the estimated baseline equilibrium and compare the impact of three different policy tools. First, we consider a standard R&D subsidy, as described above, which lowers firms' total R&D wage bill, both for variable and fixed costs. Second, we explore a spatial expansion subsidy that is proportional to firms' expenditure on fixed plant costs, $w_{ni} f_{\omega n}$. This subsidy targets the inefficiency identified in Proposition 6: when $\beta < \eta$, the planner's optimal ratio of expansion expenditure to R&D expenditure, $x_n(z)/y_n(z)$, exceeds the equilibrium ratio. A subsidy proportional to fixed plant costs $w_{ni} f_{\omega n}$ directly increases the return to expansion spending, moving the allocation of innovation labor toward the planner's optimum. A uniform version of this subsidy would achieve the planner's optimal ratio of expansion to R&D expenditure in the case of homogeneous firms and locations. However, two important aspects of the proportional expansion subsidy should be noted. First, by construction, it lowers the cost of expansion more significantly in locations with higher innovation wages. Second, it is challenging to implement, as fixed costs (and expansion costs more generally) are often unobserved, especially when they vary randomly across firms.

The third policy tool we consider is a per-market expansion subsidy, which grants firms a transfer τ for each additional market in which they open an innovative plant. While this subsidy is relatively simple to implement, it is not designed to replicate the planner's allocation. Although it encourages greater expansion, it also reallocates resources to less productive firms and markets.

To compare these policies, we choose the scale of each policy such that its total fiscal cost equals 0.1% of GDP, aligning with the estimated cost of U.S. innovation incentives to private firms such as the R&D tax credit (Brosy, 2023). All subsidies are financed by a lump-sum tax on households. We then compute the new steady-state balanced growth path under each policy. Across steady states, the distribution of idiosyncratic preference shocks ε_{ns} is held fixed; agents re-optimize their location and sector choices given the new equilibrium wages and rents, but do not draw new preference shocks.

Welfare metric. Our benchmark welfare metric is the expression for ex-ante expected utility from Equation (14).[28] This metric captures both the aggregate and distributional effects of each policy, reflecting our specification of agents' sectoral and locational preferences. For example, holding other factors constant, welfare may decline if a policy increases the dispersion of real income across space. To clarify these channels, we decompose the welfare metric into two components: one capturing aggregate real income gains, and the other reflecting distributional implications:

$$U = \underbrace{\frac{1}{\rho}\log\frac{\Upsilon}{R^{\varpi}} + \frac{g}{\rho^2}}_{\text{PDV of aggregate real income}} + \underbrace{\frac{1}{\rho}\log\left(\sum_{n=1}^{N}\sum_{s\in\{i,p\}} B_{ns}\left(S_{ns}^{\Upsilon}/S_{ns}^{R}\right)^{\epsilon}\right)^{\frac{1}{\epsilon}}}_{\text{Distributional effects}}, \tag{25}$$

where Υ is total de-trended aggregate household income,[29] and $R \equiv \left(\sum_{n',s'} L_{n's'}R_{n'}^{\varpi}\right)^{\frac{1}{\varpi}}$ is the aggregate price index of local housing services (so that R^{ϖ} serves as the aggregate consumer price index). The term $S_{ns}^{\Upsilon} \equiv \frac{L_{ns}\Upsilon_{ns}}{\sum_{n',s'} L_{n's'}\Upsilon_{n's'}}$ denotes the share of agents in location n, sector s in aggregate income, while $S_{ns}^{R} \equiv \frac{L_{ns}R_{n}^{\varpi}}{\sum_{n',s'} L_{n's'}R_{n'}^{\varpi}}$ captures the relative cost of housing faced by agents in location n, sector s.

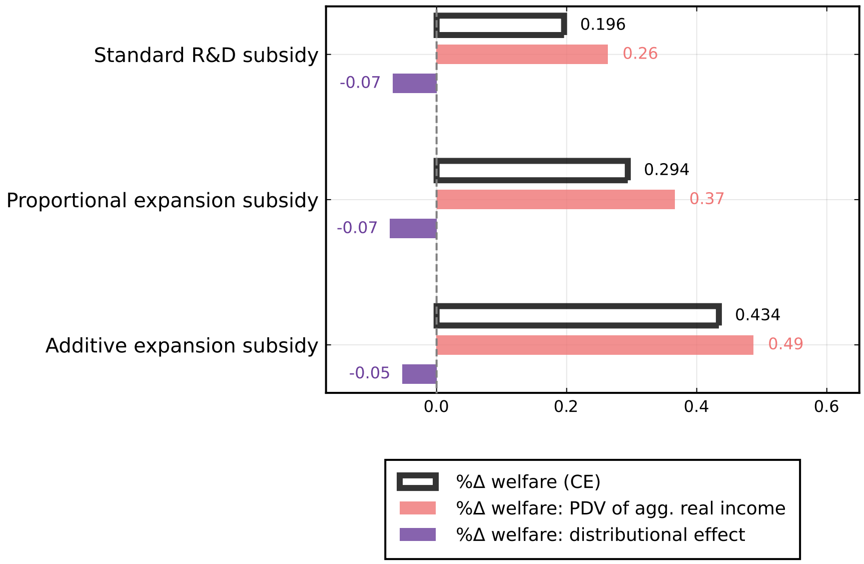
The first term in the welfare expression corresponds to the present discounted value of aggregate real income. The second term captures distributional considerations: it is higher when the income share S_{ns}^{Υ} is relatively large or the cost of living S_{ns}^{R} is relatively low in location-sector pairs that agents tend to prefer, as indicated by high amenity values B_{ns}. It is also higher when, other things equal, the dispersion in amenities-weighted real income across space and sectors is high.

Results. Expansion subsidies yield higher gains per unit of fiscal cost than the standard R&D subsidy. Figure 7 reports the aggregate gains from each of our considered policies. An R&D

[28] We adjust this equation by deducting the lump-sum tax from each agent's income.

[29] Specifically, $\Upsilon \equiv \varsigma \sum_{n,s} L_{ns}w_{ns} - T$, where T denotes the total lump-sum tax and ς is the ratio of total income (including redistributed rents and profits) to the aggregate wage bill. The decomposition follows from factoring the argument of Equation (14) into an aggregate component Υ/R^{ϖ} and location-sector-specific shares $S_{ns}^{\Upsilon}/S_{ns}^{R}$.

Figure 7: **Subsidizing R&D vs. spatial expansion – aggregate implications**



Note: This figure reports the percentage change in consumption-equivalent welfare and its components, as defined in Equation (25), under each of the three policy interventions discussed in Section 5. The white-outlined bars indicate the total change in consumption-equivalent welfare relative to the benchmark equilibrium. The red bars represent the change in the present discounted value of aggregate real income, corresponding to the first term in Equation (25). The blue bars capture the distributional welfare effect, associated with the second term in Equation (25).

subsidy equivalent to 0.1% of GDP raises consumption-equivalent welfare by 0.20%. This increase is driven by a 0.26% rise in the present discounted value of aggregate real income. However, the distributional effect is negative, equivalent to a 0.07% welfare loss. This reflects the regressive nature of R&D subsidies, which tend to raise real income in the innovation sector and in relatively richer markets ex ante. The gains from the proportional expansion subsidy are higher by approximately 0.10 percentage points, or about 50% more than the gains from the textbook R&D subsidy. Like the R&D subsidy, it reallocates more resources toward innovation. In addition, it expands the spatial scope of innovative firms, increasing knowledge spillovers in the economy. These larger gains are consistent with the results from Proposition 6 and our estimated values of η and β, and are quantitatively meaningful.

Turning to the additive expansion subsidy, it generates even higher aggregate gains and smaller negative distributional effects than the proportional expansion subsidy. The smaller distributional effect is to be expected, as this subsidy incentivizes relatively more expansion, and thus increases demand for innovation labor, in ex-ante smaller and poorer regions. Its larger aggregate gains reflect the fact that, at small policy scales, a flat per-market transfer is particularly effective at shifting the ratio of expansion to R&D expenditure toward the planner's optimum. However, this advantage diminishes at larger scales, as we elaborate in the discussion of Figure 9.

The ranking of policies is robust to shutting down firm entry and labor reallocation, though the magnitudes change. To isolate the contribution of each margin, we repeat the policy counterfactuals in restricted versions of the model that shut down each of these adjustment mechanisms. Results are shown in Figure 8. Each panel replicates the analysis from Figure 7 under a different restriction. The left panel considers a version of the model in which we shut down free entry, holding the mass of firms fixed at its baseline equilibrium level. The middle panel shuts down spatial labor reallocation, keeping total employment in each market fixed at its baseline level. The right panel shuts down both entry and labor reallocation, across sectors and across space.

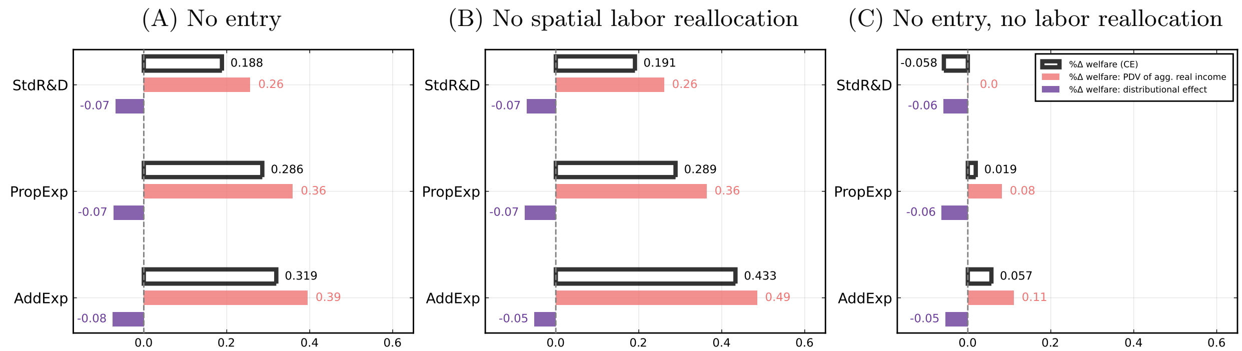
Shutting down the entry of new firms (left panel of Figure 8) and fixing the mass of firms at its baseline equilibrium level leads to smaller increases in the present discounted value of aggregate real income across all policies. This result is intuitive, as the restriction limits the reallocation of resources into innovation. The reduction in gains is notable under the additive expansion subsidy, whose welfare gain falls from 0.43% to 0.32%, narrowing its advantage over the proportional subsidy. This is because the flat per-market transfer raises expected profits for marginal entrants more than the proportional subsidy does, so shutting down entry removes a larger share of its gains. Nevertheless, the additive expansion subsidy remains the best-performing policy even when entry is shut down, and the proportional expansion subsidy continues to generate approximately 0.10 percentage points higher welfare gains than the standard R&D subsidy. This advantage arises because the gains from broader spatial expansion – via enhanced knowledge spillovers across innovative firms – remain in place even when firm entry is restricted.

Restricting labor reallocation across space (middle panel of Figure 8) has minimal impact on our baseline results, and the ranking of the different policy interventions remains unchanged. Recall that in our quantitative model, labor may be misallocated across space in the baseline equilibrium, and any policy could potentially improve or worsen this margin of efficiency. Nevertheless, we find that the welfare effects of all policies through changes in the spatial allocation of labor are quantitatively small. In particular, these effects are marginal compared to the welfare gains from reallocating resources toward innovation, and from expanding the spatial scope of innovative firms.

Finally, in the version of the model in which we shut down both firm entry and labor reallocation across space and sectors (right panel of Figure 8), the standard R&D subsidy generates a welfare *loss* of 0.06%, since resources in the innovation sector are held fixed and the subsidy raises innovation wages relative to production wages, increasing the dispersion of real income across sectors. Nevertheless, the proportional expansion subsidy still raises the present discounted value of aggregate real income by 0.08%, due to its positive effect on local knowledge spillovers across innovative firms. The additive expansion subsidy raises aggregate real income by even more (0.11%), and both expansion subsidies generate small positive welfare gains in consumption-equivalent terms (0.02% and 0.06%, respectively). Note that under all policies, the negative distributional effects remain similar to those in the baseline equilibrium. However, a planner could achieve a Pareto improvement by combining the expansion subsidies with a redistribution policy.

To shed further light on the ranking of the different policies and how it varies with the scale of the intervention, Figure 9 plots the percentage change in the present discounted value of aggregate real income against the overall fiscal cost of each policy (as a share of GDP) for the three restricted

Figure 8: **Welfare implications in different models**



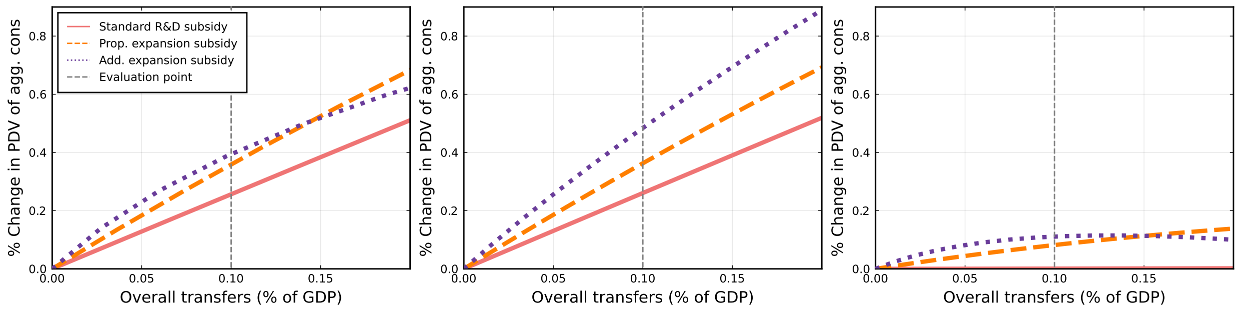
Note: This figure repeats the results from Figure 7 for restricted versions of the model. Each panel reports the percentage change in consumption-equivalent welfare and its components, as defined in Equation (25), under each of the three policy interventions discussed in Section 5. The white-outlined bars indicate the total change in consumption-equivalent welfare relative to the benchmark equilibrium. The red bars represent the change in the present discounted value of aggregate real income, corresponding to the first term in Equation (25). The blue bars capture the distributional welfare effect, associated with the second term in Equation (25). The left panel considers a version of our baseline model in which we shut down free entry, holding the mass of firms fixed at its baseline equilibrium level. The middle panel shuts down spatial labor reallocation, keeping total employment in each market fixed at its baseline level. The right panel shuts down both entry and labor reallocation, across sectors and across space.

model variants, ranging the fiscal cost from zero to 0.2% of GDP – twice our benchmark of 0.1%. These plots complement the point estimates presented in Figure 8 by showing how aggregate gains vary with the scale of the policy intervention. Several patterns emerge. First, all three policy curves eventually bend downward and reach a maximum, reflecting diminishing and ultimately negative marginal returns to each subsidy. Second, the additive expansion subsidy reaches its maximum at a smaller fiscal cost than the proportional subsidy, consistent with the overshooting mechanism discussed below. Third, the maximum attained by the proportional expansion subsidy is always above that of the additive subsidy. Fourth, the additive expansion subsidy outperforms the proportional subsidy only at small policy scales – near and below our benchmark of 0.1% of GDP – while the proportional subsidy dominates at larger scales. These patterns are robust across model variants, including those that shut down firm entry, confirming that they are not driven by differential effects on entry.

The explanation lies in how each subsidy affects the ratio of expansion expenditure to R&D expenditure, $x_n(z)/y_n(z)$. As established in Proposition 6, the equilibrium features too little expansion relative to R&D when $\beta < \eta$, and the planner would prefer a higher x/y ratio. At small scales, the per-market transfer is highly effective at shifting this ratio: because it is a flat subsidy per additional market, it has a disproportionately large effect on marginal expansion decisions – precisely the cases where the gap between the equilibrium and optimal x/y ratio matters most. In contrast, at large scales, the additive subsidy overshoots the planner's optimal x/y ratio. Because the per-market transfer does not scale with firms' productivity or with the profitability of individual markets, it pushes low-productivity firms to expand into markets where the knowledge-spillover benefits do not justify the resource cost. This generates misallocation both across firms – as low-z

Figure 9: **Comparison of policies at different scales**

(A) No entry (B) No spatial labor reallocation (C) No entry, no labor reallocation



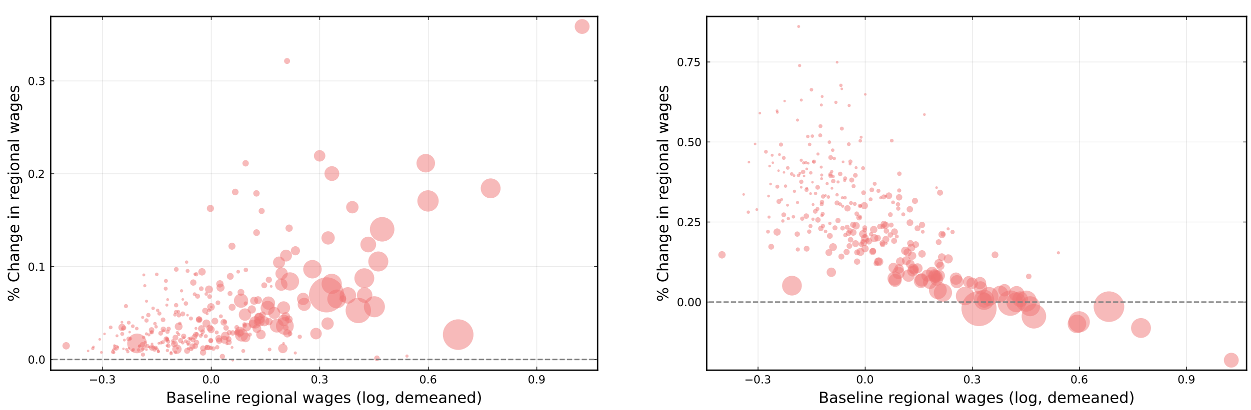
Note: This figure plots the percentage change in the present discounted value of aggregate real income against the overall fiscal cost of each policy (as a share of GDP). Each panel corresponds to a restricted version of the model: the left panel shuts down free entry, the middle panel shuts down spatial labor reallocation, and the right panel shuts down entry and labor reallocation across sectors and across space. The solid red line represents the standard R&D subsidy, the dashed orange line represents the proportional expansion subsidy, and the dotted blue line represents the additive expansion subsidy.

firms expand excessively relative to high-z firms – and across regions. The proportional expansion subsidy is targeting the ratio x/y directly, and thus at its optimum it can achieve higher gains than the additive subsidy.

5.2 Regional implications

Figure 10: **Subsidizing R&D vs. spatial expansion – regional implications**

(A) Standard R&D subsidy (B) Additive expansion subsidy



Note: This figure shows the regional implications of the subsidies considered in Section 5. The left panel plots the percentage change in regional wages from implementing the standard R&D subsidy, relative to the baseline equilibrium, against the log of average regional wages in the baseline. Each circle represents a local labor market area, weighted by population. The right panel replicates the left panel for the case of the additive expansion subsidy.

Beyond the aggregate gains, the model also predicts significant spatial heterogeneity in response to the above policies. A useful benchmark for comparison is again the standard R&D subsidy, which tends to exacerbate spatial inequality. This occurs because the subsidy primarily benefits locations that already specialize in innovation, which are characterized by higher wages in the baseline equilibrium. To illustrate this, the left panel of Figure 10 shows the percentage change in regional wages between the baseline equilibrium and the scenario with an R&D subsidy, plotted against the baseline (demeaned) log of regional wages. Each circle represents a U.S. labor market area, with the circle size reflecting the area's population. The results show that wage gains are higher in initially high-wage regions. A similar pattern emerges with the proportional expansion subsidy: although it generates positive aggregate gains, these benefits are concentrated in higher-wage markets, which experience the greatest reduction in the cost of expansion.

In contrast, the per-market expansion subsidy generates a greater incentive to expand into less profitable markets. The right panel of Figure 10 demonstrates this point. Locations with ex-ante lower wages see the largest wage increases – up to 0.75% – following the imposition of the subsidy, which results from higher growth of local innovation activity. Therefore, this variant of subsidy is able to both raise growth and reduce spatial disparities in the baseline model.

Taken together, the aggregate and regional results point to a consistent conclusion: subsidies designed to stimulate spatial expansion generally outperform standard R&D subsidies, both in terms of aggregate welfare and in their distributional implications. While the proportional expansion subsidy – which targets directly the ratio of expansion to R&D expenditure from Proposition 6 – achieves the highest gains at larger policy scales, the additive expansion subsidy is particularly effective at scales similar to existing R&D support programs and is the only policy that combines aggregate gains with spatial convergence.

6 Conclusion

When knowledge spillovers are local, the geographic scope of firms influences not only their own innovation output but also that of surrounding firms, with aggregate implications for growth and spatial inequality. We examine this mechanism through the lens of an endogenous growth model featuring multi-location innovative firms and localized spillovers. Our analysis shows that, when the diminishing returns in firms' spillovers are strong enough, firms may operate in too few markets relative to the social optimum.

Empirically, we document that firms that operate R&D facilities across multiple local markets account for most of U.S. innovation output. We also show that firms' spatial expansion into a new market increases knowledge spillovers to nearby firms, as measured by patent citations and patenting efficiency, and that these spillovers are larger when the expanding firm has a larger local footprint. Estimating the model using comprehensive data on R&D locations, patents, and citation networks, we find parameter values consistent with under-expansion of innovative firms across locations. Counterfactual simulations suggest that policies that encourage further geographical expansion of R&D activity with modest financial incentives could increase growth, outperforming traditional R&D subsidies, and, depending on implementation, reduce spatial inequality.

References

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

Akcigit, Ufuk and William R. Kerr, "Growth through Heterogeneous Innovations," *Journal of Political Economy*, 2018, *126* (4), 1374–1443.

_ , John Grigsby, Tom Nicholas, and Stefanie Stantcheva, "Taxation and Innovation in the Twentieth Century*," *The Quarterly Journal of Economics*, February 2022, *137* (1), 329–385.

_ , Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi, "Dancing with the Stars: Innovation Through Interactions," Working Paper 24466, National Bureau of Economic Research March 2018.

Argente, David, Sara Moreira, Ezra Oberfield, and Venky Venkateswaran, "Scalable Expertise," Manuscript 2020.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager, "Synthetic Difference-in-Differences," *American Economic Review*, 2021, *111* (12), 4088–4118.

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

Audretsch, David B and Maryann P Feldman, "R&D spillovers and the geography of innovation and production," *The American economic review*, 1996, *86* (3), 630–640.

Ayerst, Stephen, Faisal Ibrahim, Gaelan MacKenzie, and Swapnika Rachapalli, "Trade and diffusion of embodied technology: An empirical analysis," *Journal of Monetary Economics*, 2023, *137*, 128–145.

Barnatchez, Keith, Leland D. Crane, and Ryan A. Decker, "An Assessment of the National Establishment Time Series (NETS) Database," *Federal Reserve Board of Governors Working Paper*, 2017.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein, "Synthetic Controls with Staggered Adoption," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2022, *84* (2), 351–381.

Bergé, Laurent, "Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm," Technical Report 13 2018.

Berkes, Enrico and Ruben Gaetani, "The Geography of Unconventional Innovation," *The Economic Journal*, 09 2020, *131* (636), 1466–1514.

_ , _ , **and Martí Mestieri**, "Technological Waves, Knowledge Diffusion, and Local Growth," *Journal of Political Economy Macroeconomics*, 2025, *3* (1), 75–121.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen, "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 2013, *81* (4), 1347–1393.

Brosy, Thomas, "Understanding R&D Tax Breaks and Reform Options | Tax Policy Center," June 2023.

Buera, Francisco J. and Ezra Oberfield, "The Global Diffusion of Ideas," *Econometrica*, 2020, *88* (1), 83–114.

Burchardi, Konrad B., Thomas Chaney, Tarek Alexander Hassan, Lisa Tarquinio, and Stephen J. Terry, "Immigration, Innovation, and Growth," *American Economic Review*, 2026, *116* (3), 828–861.

Cai, Jie, Nan Li, and Ana Maria Santacreu, "Knowledge Diffusion, Trade, and Innovation across Countries and Sectors," *American Economic Journal: Macroeconomics*, 2022, *14* (1), 104–45.

Cai, Sheng, Lorenzo Caliendo, Fernando Parro, and Wei Xiang, "Mechanics of Spatial Growth," Working Paper 30579, National Bureau of Economic Research 2025.

Carlino, Gerald and William R. Kerr, "Chapter 6 - Agglomeration and Innovation," in "Handbook of Regional and Urban Economics," Vol. 5, Elsevier, 2015, pp. 349–404.

Cen, Ling, Edward L. Maydew, Liandong Zhang, and Luo Zuo, "Customer–supplier relationships and corporate tax avoidance," *Journal of Financial Economics*, February 2017, *123* (2), 377–394.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer, "The Effect of Minimum Wages on Low-Wage Jobs," *The Quarterly Journal of Economics*, 2019, *134* (3), 1405–1454.

Chatterji, Aaron, Edward Glaeser, and William Kerr, "Clusters of Entrepreneurship and Innovation," *Innovation Policy and the Economy*, 2014, *14*, 129–166.

Cohen, Lauren and Andrea Frazzini, "Economic Links and Predictable Returns," *The Journal of Finance*, 2008, *63* (4), 1977–2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2008.01379.x.

Comin, Diego A, Danial Lashkari, and Martí Mestieri, "The Structural Transformation of Innovation," Technical Report, National Bureau of Economic Research 2025.

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin, "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects," *The Stata Journal*, 2020, *20* (1), 95–115.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade theory with numbers: Quantifying the consequences of globalization," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 197–261.

Crews, Levi Garrett, "A dynamic spatial knowledge economy." PhD dissertation, The University of Chicago 2023.

Davis, Morris A and François Ortalo-Magné, "Household expenditures, wages, rents," *Review of Economic Dynamics*, 2011, *14* (2), 248–261.

Davis, Steven J., Jonathan Haltiwanger, and Scott Schuh, *Job Creation and Destruction*, MIT Press, January 1998.

Dixit, Avinash K and Joseph E Stiglitz, "Monopolistic competition and optimum product diversity," *The American economic review*, 1977, *67* (3), 297–308.

Eckert, Fabian and Michael Peters, "Spatial Structural Change," Working Paper 30489, National Bureau of Economic Research 2025.

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

_ **and** _ , "Place-Based Policies: Lessons from Theory," Working Paper 33517, National Bureau of Economic Research February 2025.

Fort, Teresa C, Wolfgang Keller, Peter K Schott, Stephen Yeaple, and Nikolas Zolas, "Colocation of Production and Innovation: Evidence from the United States," *Working Paper*, 2020.

Galle, Simon, Andrés Rodríguez-Clare, and Moises Yi, "Slicing the pie: Quantifying the aggregate and distributional effects of trade," *The Review of Economic Studies*, 2023, *90* (1), 331–375.

Gardner, John, "Two-stage differences in differences," Technical Report, arXiv July 2022. arXiv:2207.05943.

Garetto, Stefania, "Input Sourcing and Multinational Production," *American Economic Journal: Macroeconomics*, April 2013, *5* (2), 118–51.

Giroud, Xavier, Ernest Liu, and Holger Mueller, "Innovation Spillovers across U.S. Tech Clusters," *Journal of Financial Economics*, 2026.

_ **, Simone Lenzu, Quinn Maingi, and Holger Mueller**, "Propagation and Amplification of Local Productivity Spillovers," *Econometrica*, 2024, *92* (5), 1589–1619.

Glaeser, Edward L and Joshua D Gottlieb, "The economics of place-making policies," Technical Report, National Bureau of Economic Research 2008.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings," *Journal of political economy*, 2010, *118* (3), 536–598.

Griffith, Rachel, Sokbae Lee, and John Van Reenen, "Is distance dying at last? Falling home bias in fixed-effects models of patent citations," *Quantitative Economics*, 2011, *2* (2), 211–249.

Grossman, Gene M. and Elhanan Helpman, "Quality Ladders in the Theory of Growth," *The Review of Economic Studies*, 1991, *58* (1), 43–61.

Helpman, Elhanan, "A simple theory of international trade with multinational corporations," *Journal of political economy*, 1984, *92* (3), 451–471.

Hsieh, Chang-Tai, Peter J. Klenow, and Ishan Nath, "A Global View of Creative Destruction," *Journal of Political Economy Macroeconomics*, 2023, *1* (2), 243–275.

Hughes, Ryan, Charles deGrazia, and Julian Kolev, "Technical Documentation for Matching Patents and Trademarks to the 2017 National Establishment Time Series Database," https://www.uspto.gov/sites/default/files/documents/oce-wp-ip-to-nets.pdf 2021.

Jacobs, Jane, *The economy of cities*, Vintage, 1969.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *The Quarterly Journal of Economics*, 1993, *108* (3), 577–598. Publisher: Oxford University Press.

Keller, Wolfgang and Stephen Ross Yeaple, "The Gravity of Knowledge," *American Economic Review*, June 2013, *103* (4), 1414–44.

Kerr, William R., "Centralization and Organization Reproduction: Ethnic Innovation in R&D Centers and Satellite Locations," *Management Science*, 2020, *66* (10), 4470–4492.

_ **and Frederic Robert-Nicoud**, "Tech Clusters," *Journal of Economic Perspectives*, 2020, *34* (3), 50–76.

Kleinman, Benny, "Wage Inequality and the Spatial Expansion of Firms," *Working Paper*, 2023.

Klette, Tor Jakob and Samuel Kortum, "Innovating Firms and Aggregate Innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

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

Lhuillier, Hugo, "Should I Stay or Should I Grow?," 2023. Working Paper.

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

Lucas, Robert E. and Benjamin Moll, “Knowledge Growth and the Allocation of Time,” *Journal of Political Economy*, 2014, *122* (1), 1–51.

Martellini, Paolo, “Local labor markets and aggregate productivity,” Technical Report, Working paper 2022.

Matray, Adrien, “The local innovation spillovers of listed firms,” *Journal of Financial Economics*, 2021, *141* (2), 395–412.

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

Oberfield, Ezra, Esteban Rossi-Hansberg, Pierre-Daniel Sarte, and Nicholas Trachter, “Plants in space,” *Journal of Political Economy*, 2024, *132* (3), 000–000.

Pakes, Ariel and Zvi Griliches, “Patents and R and D at the Firm Level: A First Look,” Technical Report 0561, National Bureau of Economic Research 1980.

Pauly, Stefan and Fernando Stipanicic, “The creation and diffusion of knowledge: Evidence from the Jet Age,” CEPREMAP Working Papers (Docweb) 2112, CEPREMAP November 2025.

Peri, Giovanni, “Determinants of Knowledge Flows and Their Effect on Innovation,” *The Review of Economics and Statistics*, 2005, *87* (2), 308–322.

Perla, Jesse and Christopher Tonetti, “Equilibrium Imitation and Growth,” *Journal of Political Economy*, 2014, *122* (1), 52–76.

_ , _ , and Michael E. Waugh, “Equilibrium Technology Diffusion, Trade, and Growth,” *American Economic Review*, 2021, *111* (1), 73–128.

Porter, Michael E. and Scott Stern, “Measuring the ”Ideas” Production Function: Evidence from International Patent Output,” *NBER Working Papers*, September 2000. Number: 7891.

Prato, Marta, “The Global Race for Talent: Brain Drain, Knowledge Transfer, and Growth,” *The Quarterly Journal of Economics*, 2025, *140* (1), 165–238.

Ramondo, Natalia and Andrés Rodríguez-Clare, “Trade, multinational production, and the gains from openness,” *Journal of Political Economy*, 2013, *121* (2), 273–322.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman, “Cognitive Hubs and Spatial Redistribution,” *American Economic Journal: Macroeconomics*, 2025.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023, *235* (2), 2218–2244.

Sampson, Thomas, "Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth," *The Quarterly Journal of Economics*, 2015, *131* (1), 315–380.

Santacreu, Ana Maria, "Innovation, diffusion, and trade: Theory and measurement," *Journal of Monetary Economics*, 2015, *75*, 1–20.

Scherer, F. M and Dietmar Harhoff, "Technology policy for a world of skew-distributed outcomes," *Research Policy*, April 2000, *29* (4), 559–566.

Terry, Stephen J, Toni M Whited, and Anastasia A Zakolyukina, "Information versus Investment," *The Review of Financial Studies*, 2023, *36* (3), 1148–1191.

Tintelnot, Felix, "Global production with export platforms," *The Quarterly Journal of Economics*, 2017, *132* (1), 157–209.

Wasi, Nada and Aaron Flaaen, "Record Linkage Using Stata: Preprocessing, Linking, and Reviewing Utilities," *The Stata Journal*, 2015, *15* (3), 672–697. Publisher: SAGE Publications.

A Proofs and derivations

A.1 Equilibrium conditions

We summarize the conditions for a balanced growth path equilibrium, which consists of an allocation of production and innovation workers for all markets, production and innovation wages in all markets, quantity and price of the final and intermediate goods, a mass of innovative firms, and the value from new innovations, that satisfy:

1. Labor market clearing in the production sector

$$w_{np} = \frac{1}{\gamma} Z_n^{\frac{\sigma-1}{\sigma}} \left(\frac{\bar{L}_n}{Y} \right)^{-\frac{1}{\sigma}}. \tag{26}$$

2. Labor market clearing in the innovation sector

$$\bar{H}_n = M \int_z \chi_n(z) \left(\bar{f}_n(z) + \ell_n(z) \right) d\Psi(z), \tag{27}$$

   where we solved for $\ell_n(z)$, $\chi_n(z)$, and $\bar{f}_n(z)$ above in terms of equilibrium objects in Section 2.2.2.

3. Regional innovation productivity is given by

$$K_n = \bar{K}_n \left(M \int_z \chi_n(z)\, z \ell_n(z)^{\beta} d\Psi(z) \right)^{\alpha}. \tag{28}$$

4. Market clearing in the final good market

$$Y = M f_e + (r-g)V + \sum_{n=1}^{N} w_{ni} \bar{H}_n + \sum_{n=1}^{N} w_{np} \bar{L}_n. \tag{29}$$

5. The value of new innovations is given by Equation (6).

6. Free entry into innovation satisfies Equation (8).

7. Aggregate productivity $\mathcal{A}_t$, wages, consumption, profits, and the value of new innovations grow at a constant rate g, given by Equation (10).

A.2 Proof of Lemma 1 and Proposition 2.

Consider the problem of an innovative firm ω described in equation (7).

Conditional on being active in market n, the firm's innovative employment decision for such market solves the following problem:

$$\max_{\ell_{n\omega t}} \lambda_{\omega t} V_t - \left[w_{nit} \left(\ell_{n\omega t} + f_{n\omega t} \right) \right]$$

$$\text{s.t.} \quad \lambda_{\omega t} = \lambda_0 \sum_{n=1}^{N} K_{nt} z_\omega (\ell_{n\omega t})^\eta$$

The first order condition of this problem is the following:

$$\eta \lambda_0 K_{nt} z_\omega (\ell_{n\omega t})^{\eta-1} V_t = w_{nit}.$$

Rearranging this expression, we obtain:

$$\ell_{n\omega t} = \left(z_\omega \frac{\eta K_{nt} \lambda_0 V_t}{w_{nit}} \right)^{\frac{1}{1-\eta}}.$$

Given the optimal innovative employment conditional on activity in market n, a firm will choose to enter such market if the additional idea arrival rate obtained from an R&D establishment in that location is higher than the total fixed and variable labor costs, as described by the following problem:

$$\max\{\lambda_0 K_{nt} z_\omega (\ell_{n\omega t})^\eta V_t - \left[w_{nit} \left(\ell_{n\omega t} + f_{n\omega t} \right) \right], 0\}$$

$$\text{s.t.} \quad \ell_{n\omega t} = \left(z_\omega \frac{\eta K_{nt} \lambda_0 V_t}{w_{nit}} \right)^{\frac{1}{1-\eta}}.$$

The solution to this problem is a threshold rule, which indicates that the firm will enter market n if the draw for $f_{n\omega t}$ is low enough :

$$f_{n\omega t} < \left(\frac{1-\eta}{\eta} \right) \ell_{n\omega t}.$$

As a result, the probability that a firm ω chooses to be active in market n, before observing the fixed cost draw, is:

$$\chi_{n\omega t} = \Pr \left(f_{n\omega t} < \left(\frac{1-\eta}{\eta} \right) \ell_{n\omega t} \right) = f_{max}^{-\theta} \left(\left(\frac{1-\eta}{\eta} \right) \ell_{n\omega t} \right)^\theta,$$

where the second equality follows from the assumption that the fixed cost distribution is inverse Pareto.

The expected amount of labor that a firm ω hires in location n to pay the fixed costs, conditional on being active in n, is:

$$\bar{f}_{n\omega t} = \mathbb{E}\left[f \middle| f < \left(\frac{1-\eta}{\eta}\right)\ell_{n\omega t}\right] = \int_0^{\left(\frac{1-\eta}{\eta}\right)\ell_{n\omega t}} f \frac{\theta f_{max}^{-\theta} f^{\theta-1}}{f_{max}^{-\theta}\left(\frac{1-\eta}{\eta}\ell_{n\omega t}\right)^{\theta}} df = \frac{\theta}{\theta+1}\left(\frac{1-\eta}{\eta}\ell_{n\omega t}\right). \quad (30)$$

We note that the optimal innovative employment conditional on activity, the probability of entering a market, and the expected fixed cost are the same for all firms with productivity z, so we suppress the index ω and write $\ell_{nt}(z) = \ell_{n\omega t}$, $\chi_{nt}(z) = \chi_{n\omega t}$, and $\bar{f}_{nt}(z) = \bar{f}_{n\omega t}$ for all ω such that $z_\omega = z$.

Finally, from Equation (30), we obtain that the ratio of the expected amount of region-n innovation labor allocated to expansion relative to R&D activities for type-z firms is:

$$\frac{x_n(z)}{y_n(z)} = \frac{\chi_n(z)\bar{f}_n(z)}{\chi_n(z)\ell_n(z)} = \frac{1-\eta}{\eta}\frac{\theta}{\theta+1}.$$

A.3 Proof of Proposition 3.

The innovation output of a firm with productivity z in market n, conditional on activity, is $\lambda_n(z) = \lambda_0 K_n z \ell_n(z)^\eta$. From the expressions in Section A.2, $\ell_n(z) = z^{\frac{1}{1-\eta}}\Omega_n$ where $\Omega_n \equiv \left(\frac{\eta K_n \lambda_0 V}{w_{ni}}\right)^{\frac{1}{1-\eta}}$. Substituting:

$$\lambda_n(z) = \lambda_0 K_n z^{\frac{1}{1-\eta}}\Omega_n^\eta.$$

From Section A.4, innovation labor market clearing implies $\Omega_n = C_\Omega H_{RD,n}^{\frac{1}{1+\theta}}$, where C_Ω depends on M, f_{max}, and moments of Ψ but not on n or z. Moreover, $K_n = \bar{K}_n \bar{z}_\beta^\alpha M_n^{\alpha(1-\beta)} H_{RD,n}^{\alpha\beta}$ and $M_n = \tilde{c}\, H_{RD,n}^{\frac{\theta}{1+\theta}}$. Substituting:

$$K_n = \bar{K}_n \bar{z}_\beta^\alpha \tilde{c}^{\alpha(1-\beta)} H_{RD,n}^{\frac{\alpha\theta(1-\beta)}{1+\theta}+\alpha\beta} = \bar{K}_n \bar{z}_\beta^\alpha \tilde{c}^{\alpha(1-\beta)} H_{RD,n}^{\frac{\alpha(\theta+\beta)}{1+\theta}}.$$

Therefore:

$$\begin{aligned}\lambda_n(z) &= \lambda_0 \bar{K}_n \bar{z}_\beta^\alpha \tilde{c}^{\alpha(1-\beta)} H_{RD,n}^{\frac{\alpha(\theta+\beta)}{1+\theta}} z^{\frac{1}{1-\eta}} C_\Omega^\eta H_{RD,n}^{\frac{\eta}{1+\theta}} \\ &= \tilde{C}_\lambda \bar{K}_n z^{\frac{1}{1-\eta}} H_{RD,n}^{\frac{\eta+\alpha(\theta+\beta)}{1+\theta}},\end{aligned}$$

where $\tilde{C}_\lambda \equiv \lambda_0 \bar{z}_\beta^\alpha \tilde{c}^{\alpha(1-\beta)} C_\Omega^\eta > 0$. Since $H_{RD,n} = \frac{\eta(1+\theta)}{\eta+\theta} H_n$, where H_n is total innovation employment in market n, we can absorb the proportionality constant into $\tilde{C}_\lambda$ and write:

$$\lambda_n(z) = \tilde{C}_\lambda \bar{K}_n z^{\frac{1}{1-\eta}} H_n^{\frac{\eta+\alpha(\theta+\beta)}{1+\theta}}.$$

A.4 Proof of Proposition 4.

Part (i): Regional innovation output.

The result holds at any time t; we suppress time subscripts in the proof for readability. Regional innovation output, λ_n, is given by the following expression:

$$\lambda_n = MK_n \int_z \chi_n(z) z \left(\ell_n(z)\right)^\eta d\Psi(z).$$

Define the mass of R&D firms of type z in location n as $M_n(z) \equiv M\chi_n(z)$ and the mass of R&D labor in firms of type z in location n as $H_{RD,n}(z) \equiv M\chi_n(z)\ell_n(z)$. In addition, define the regional mass of R&D firms and R&D labor as $M_n \equiv \int_z M_n(z)\, d\Psi(z)$ and $H_{RD,n} \equiv \int_z H_{RD,n}(z)\, d\Psi(z)$ respectively. Note that, using the expressions for $\chi_n(z)$ and $\ell_n(z)$ from Section A.2, we can write:

$$M_n(z) = \frac{z^{\frac{\theta}{1-\eta}}}{\int_z z^{\frac{\theta}{1-\eta}} d\Psi(z)} M_n$$
$$H_{RD,n}(z) = \frac{z^{\frac{1+\theta}{1-\eta}}}{\int_z z^{\frac{1+\theta}{1-\eta}} d\Psi(z)} H_{RD,n}.$$

Using these expressions, we note that, for some constant $a > 0$, we can solve the following expression:

$$\begin{aligned}
\mathcal{X}_{a,n} &= \int_z z\left(M_n(z)\right)^{1-a}\left(H_{RD,n}(z)\right)^a d\Psi(z) \\
&= \int_z z\left(\frac{z^{\frac{\theta}{1-\eta}}}{\int_z z^{\frac{\theta}{1-\eta}} d\Psi(z)} M_n\right)^{1-a}\left(\frac{z^{\frac{1+\theta}{1-\eta}}}{\int_z z^{\frac{1+\theta}{1-\eta}} d\Psi(z)} H_{RD,n}\right)^a d\Psi(z), \\
&= \bar{z}_a M_n^{1-a} H_{RD,n}^a,
\end{aligned}$$

where we define

$$\bar{z}_a \equiv \int_z z\left(\frac{z^{\frac{\theta}{1-\eta}}}{\int_z z^{\frac{\theta}{1-\eta}} d\Psi(z)}\right)^{1-a}\left(\frac{z^{\frac{1+\theta}{1-\eta}}}{\int_z z^{\frac{1+\theta}{1-\eta}} d\Psi(z)}\right)^a d\Psi(z).$$

Now we can solve for regional innovation output as follows:

$$\begin{aligned}
\lambda_n &= K_n \int_z M\chi_n(z)\, z\left(\ell_n(z)\right)^\eta d\Psi(z) \\
&= K_n \bar{z}_\eta M_n^{1-\eta} H_{RD,n}^\eta \\
&= \bar{K}_n\left(\int_z M\chi_i(z)\, z\left(\ell_i(z)\right)^\beta d\Psi(z)\right)^\alpha \bar{z}_\eta M_n^{1-\eta} H_{RD,n}^\eta \\
&= \bar{K}_n \bar{z}_\beta^\alpha \bar{z}_\eta M_n^{(1-\eta)+\alpha(1-\beta)} H_{RD,n}^{\eta+\alpha\beta}.
\end{aligned}$$

From Proposition 2, the ratio of expansion to R&D labor is constant across firms and regions:

$$\frac{x_n(z)}{y_n(z)} = \frac{1-\eta}{\eta}\frac{\theta}{\theta+1}.$$

Aggregating across firm types, total regional innovation labor satisfies $\bar{H}_n = H_{RD,n}\left(1 + \frac{1-\eta}{\eta}\frac{\theta}{\theta+1}\right)$, so that $H_{RD,n} = \frac{\eta(1+\theta)}{\eta+\theta}\bar{H}_n$. Substituting and collecting constants into $C_\lambda \equiv \bar{z}_\eta\, \bar{z}_\beta^\alpha \left(\frac{\eta(1+\theta)}{\eta+\theta}\right)^{\eta+\alpha\beta}$ yields equation (11).

We next derive the elasticity of average regional innovation output, λ_n/M_n, with respect to total regional innovation labor, H_n, which is used in Section 4.2.1. To do so, we express M_n in terms of $H_{RD,n}$.

From the equilibrium expressions for $\chi_n(z)$ and $\ell_n(z)$ in Section A.2, we have:

$$\chi_n(z) = f_{max}^{-\theta}\left(\frac{1-\eta}{\eta}\right)^{\theta}\ell_n(z)^{\theta},$$
$$\ell_n(z) = z^{\frac{1}{1-\eta}}\left(\frac{\eta K_n \lambda_0 V}{w_{ni}}\right)^{\frac{1}{1-\eta}}.$$

Define $\Omega_n \equiv \left(\frac{\eta K_n \lambda_0 V}{w_{ni}}\right)^{\frac{1}{1-\eta}}$, so that $\ell_n(z) = z^{\frac{1}{1-\eta}}\Omega_n$ and $\chi_n(z) = f_{max}^{-\theta}\left(\frac{1-\eta}{\eta}\right)^{\theta} z^{\frac{\theta}{1-\eta}}\Omega_n^\theta$. We can then write:

$$M_n = M\int_z \chi_n(z)\, d\Psi(z) = M f_{max}^{-\theta}\left(\frac{1-\eta}{\eta}\right)^{\theta}\mathbb{E}\left[z^{\frac{\theta}{1-\eta}}\right]\Omega_n^\theta,$$
$$H_{RD,n} = M\int_z \chi_n(z)\ell_n(z)\, d\Psi(z) = M f_{max}^{-\theta}\left(\frac{1-\eta}{\eta}\right)^{\theta}\mathbb{E}\left[z^{\frac{1+\theta}{1-\eta}}\right]\Omega_n^{1+\theta}.$$

Eliminating Ω_n between these two expressions, we obtain:

$$M_n = \tilde{c}\, H_{RD,n}^{\frac{\theta}{1+\theta}},$$

where $\tilde{c}$ is a positive constant that depends on M, f_{max}, and moments of the productivity distribution, but does not vary across regions.

Moreover, since $H_{RD,n} = \frac{\eta(1+\theta)}{\eta+\theta}\bar{H}_n$ as shown above, we have $\frac{\partial \log H_{RD,n}}{\partial \log \bar{H}_n} = 1$ and $\frac{\partial \log M_n}{\partial \log \bar{H}_n} = \frac{\theta}{1+\theta}$. Taking logs of equation (11) and substituting $\log M_n = \text{const} + \frac{\theta}{1+\theta}\log H_{RD,n}$:

$$\log\left(\frac{\lambda_n}{M_n}\right) = \log \bar{K}_n + \text{const} + [(1-\eta) + \alpha(1-\beta) - 1]\log M_n + [\eta + \alpha\beta]\log H_{RD,n}$$
$$= \log \bar{K}_n + \text{const} + [\alpha(1-\beta) - \eta]\,\frac{\theta}{1+\theta}\log H_{RD,n} + [\eta+\alpha\beta]\log H_{RD,n}.$$

Taking the derivative with respect to $\log H_n$, and using $\frac{\partial \log H_{RD,n}}{\partial \log H_n} = 1$ while holding $\bar{K}_n$ fixed:

$$\frac{\partial \log\left(\lambda_n/M_n\right)}{\partial \log H_n} = \left[\alpha(1-\beta)-\eta\right]\frac{\theta}{1+\theta} + \eta + \alpha\beta = \alpha\left[\frac{\theta}{1+\theta} + \frac{\beta}{1+\theta}\right] + \frac{\eta}{1+\theta}.$$

Finally, we derive the elasticity of firm-level employment to regional innovation labor, which is used in the estimation of θ in Section 4. From the expression for $H_{RD,n}$ above, we can solve for Ω_n as a function of $H_{RD,n}$:

$$\Omega_n = \left(\frac{H_{RD,n}}{M f_{max}^{-\theta}\left(\frac{1-\eta}{\eta}\right)^{\theta}\mathbb{E}\left[z^{\frac{1+\theta}{1-\eta}}\right]}\right)^{\frac{1}{1+\theta}}.$$

Substituting back into the expression for firm-level employment, $\ell_n(z) = z^{\frac{1}{1-\eta}}\Omega_n$, and taking logs:

$$\log \ell_n(z) = \frac{1}{1-\eta}\log z + \frac{1}{1+\theta}\log H_{RD,n} + \text{const},$$

where the constant absorbs terms involving M, f_{max}, and moments of the productivity distribution. Since $H_{RD,n}$ is proportional to H_n with a region-invariant factor, this implies:

$$\log \ell_n(z) = \frac{1}{1-\eta}\log z + \frac{1}{1+\theta}\log H_n + \text{const}.$$

Thus, conditional on firm productivity z and holding $\bar{K}_n$ fixed, firm-level employment is log-linear in regional innovation labor with elasticity $\frac{1}{1+\theta}$.

Part (ii): Equilibrium growth.

Consider the expression for aggregate output in equation (1). Plugging in the expression for intermediate goods production in equation (2), we can express aggregate output as:

$$\log Y_t = \int_0^1 \log A_{jt} dj + \int_0^1 \log\left(\sum_{n=1}^{N}(Z_n L_{jn})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} dj = \log(\mathcal{A}_t) + \log(\mathcal{L}^p), \tag{31}$$

where we define aggregate productivity, $\mathcal{A}_t$, such that $\log(\mathcal{A}_t) \equiv \int_0^1 \log(A_{jt})dj$ and $\log(\mathcal{L}^p) \equiv \int_0^1 \log\left[\sum_{n=1}^{N}(Z_n L_{jn})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} dj$. Note that aggregate output and aggregate productivity grow at the same rate g.

We can write the change in aggregate productivity $\mathcal{A}_t$ in a small time interval, Δt, as $\frac{\log(\mathcal{A}_{t+\Delta t})-\log(\mathcal{A}_t)}{\Delta t}$

and obtain the growth rate on the balanced growth path by letting Δt go to zero, i.e.:

$$
\begin{aligned}
g &= \lim_{\Delta t \to 0} \frac{\log(\mathcal{A}_{t+\Delta t}) - \log(\mathcal{A}_t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\int_0^1 \left(\bar{\lambda}_t \Delta t \log\left(\gamma A_{jt}\right) + \left(1 - \bar{\lambda}_t \Delta t\right) \log\left(A_{jt}\right)\right) dj - \log(\mathcal{A}_t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\bar{\lambda}_t \Delta t \log(\gamma) + \log(\mathcal{A}_t) - \log(\mathcal{A}_t)}{\Delta t} \\
&= \bar{\lambda}_t \log(\gamma).
\end{aligned}
$$

The second line is derived by noting that, in an interval of time Δt, each intermediate j is subject to being innovated upon with probability $\bar{\lambda}_t \Delta t$. As a consequence, in the aggregate, a fraction $\bar{\lambda}_t \Delta t$ of the intermediates is hit by an innovation in the time interval Δt, and the productivity of those intermediates increases by a factor γ. Notice that a constant aggregate growth rate along a BGP requires a constant rate of creative destruction, $\bar{\lambda}$.

The aggregate rate of creative destruction is derived by combining Equations (3) and (4), which delivers:

$$
\bar{\lambda} = M\lambda_0 \int_z \sum_{n=1}^{N} K_n\left(\{\chi, \ell\}\right) \chi_n\left(z\right) z \ell_n\left(z\right)^\eta d\Psi\left(z\right).
$$

A.5 Proof of Proposition 5.

Consider the growth rate presented in Equation (10). Plugging in Equation (28) and the expressions for the innovative firms optimal decisions, we obtain:

$$
\begin{aligned}
g &= \log(\gamma) M\lambda_0 \sum_{n=1}^{N} \bar{K}_n \left(M \int_{z'} \chi_n(z') z' \ell_n(z')^\beta \, d\Psi(z')\right)^\alpha \int_z \chi_n(z) z \ell_n^\eta(z) \, d\Psi(z) \\
&= \log(\gamma) M\lambda_0 \sum_{n=1}^{N} \bar{K}_n \left(M \int_{z'} z' f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^\theta \ell_n(z')^{\theta+\beta} \, d\Psi(z')\right)^\alpha \times \\
&\qquad \times \int_z z f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^\theta \ell_n^{\theta+\eta}(z) \, d\Psi(z)
\end{aligned}
$$

Plugging the firm's optimal innovative labor hiring decision we get:

$$
g = \log(\gamma) \left(M f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^\theta\right)^{\alpha+1} \left(\mathbb{E}\left(z^{\frac{\theta+\beta}{1-\eta}+1}\right)\right)^\alpha \mathbb{E}\left(z^{\frac{\theta+1}{1-\eta}}\right) \lambda_0 \sum_{n=1}^{N} \bar{K}_n \left(\frac{\eta K_n \lambda_0 V}{w_{ni}}\right)^{\frac{\theta+\eta+\alpha(\theta+\beta)}{1-\eta}}.
$$

The innovative labor market clearing delivers the following expression:

$$\bar{H}_n = M \int_z \chi_n(z) \left(\ell_n(z) + \bar{f}_n(z)\right) \psi(z) dz$$
$$= M \int_z f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^{\theta} \left(1 + \frac{\theta}{\theta+1}\frac{1-\eta}{\eta}\right) (\ell_n(z))^{1+\theta} \psi(z) dz.$$

Combining this with the firm's innovation labor decision, we get:

$$\frac{\lambda_0 \eta K_n V}{w_{ni}} = \left(\frac{\bar{H}_n}{M\mathbb{E}[z^{\frac{1+\theta}{1-\eta}}] f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^{\theta} \left(1 + \frac{\theta}{\theta+1}\frac{1-\eta}{\eta}\right)}\right)^{\frac{1-\eta}{1+\theta}}.$$

Plugging this expression into the equation for g we obtain:

$$g = \log(\gamma)\lambda_0 \left(M f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^{\theta}\right)^{\alpha+1-\frac{\theta+\eta+\alpha(\theta+\beta)}{1+\theta}} \left(\mathbb{E}\left(z^{\frac{\theta+\beta}{1-\eta}+1}\right)\right)^{\alpha} \mathbb{E}\left(z^{\frac{\theta+1}{1-\eta}}\right)^{1-\frac{\theta+\eta+\alpha(\theta+\beta)}{1+\theta}} \times$$
$$\left(1 + \frac{\theta}{\theta+1}\frac{1-\eta}{\eta}\right)^{-\frac{\theta+\eta+\alpha(\theta+\beta)}{1+\theta}} \sum_{n=1}^{N} \bar{K}_n \bar{H}_n^{\frac{\theta+\eta+\alpha(\theta+\beta)}{1+\theta}}$$

From this expression we obtain that the elasticity of the growth rate relative to f_{max} is given by:

$$\frac{\partial \log g}{\partial \log f_{max}} = \mu \left(\frac{\partial \log M}{\partial \log f_{max}} - \theta\right), \tag{32}$$

where $\mu \equiv \frac{(1-\eta)+\alpha(1-\beta)}{1+\theta} > 0$.

Sign of $\partial \log M / \partial \log f_{max}$ under free entry. We now show that free entry implies $\partial \log M/\partial \log f_{max} < 0$. Using the firm's first-order condition, $w_{ni} = \eta\lambda_0 K_n z \ell_n(z)^{\eta-1} V$, the net expected profit conditional on activity in market n for a firm of type z can be written as:

$$\lambda_0 K_n z \ell_n(z)^{\eta} V - w_{ni}\left(\ell_n(z) + \bar{f}_n(z)\right) = w_{ni}\ell_n(z)\frac{1-\eta}{\eta}\frac{1}{\theta+1}.$$

Expected profits per firm are therefore:

$$\Pi = \frac{1-\eta}{\eta(\theta+1)} \sum_{n=1}^{N} w_{ni} \int_z \chi_n(z)\ell_n(z)\, d\Psi(z) = \frac{1-\eta}{\eta(\theta+1)} \sum_{n=1}^{N} w_{ni} \frac{H_{RD,n}}{M},$$

where $H_{RD,n} = M \int_z \chi_n(z)\ell_n(z)\, d\Psi(z)$ is total R&D labor in market n. Since total innovation revenue is $\bar{\lambda}V = \sum_n w_{ni}\bar{H}_n \frac{1+\theta}{\eta+\theta} = \sum_n w_{ni}\frac{H_{RD,n}}{\eta}$ (using $\bar{H}_n = H_{RD,n}\frac{\eta+\theta}{\eta(1+\theta)}$ and the accounting identity $\bar{\lambda}V = \sum_n w_{ni}\bar{H}_n + M\Pi$), this simplifies to:

$$\Pi = \frac{1-\eta}{M(1+\theta)}\bar{\lambda}V. \tag{33}$$

On a BGP, $\bar{\lambda} = g/\log(\gamma)$ and $V = \bar{\pi}Y/(r - g + \bar{\lambda})$. Using the Euler equation $r = \rho + \phi g$ and denoting the (constant) ratio $\tilde{f}_e \equiv f_e/Y$, the free entry condition $\Pi = f_e$ yields:

$$M = \frac{(1-\eta)\bar{\pi}}{(1+\theta)\tilde{f}_e} \cdot \frac{g}{D(g)}, \tag{34}$$

where $D(g) \equiv \rho\log(\gamma) + g\left[(\phi-1)\log(\gamma)+1\right] > 0$.

The growth expression derived above can be written compactly as $g = A\, M^{\mu}\, f_{max}^{-\theta\mu}$, where $A > 0$ collects terms that depend only on exogenous objects ($\bar{K}_n$, $\bar{H}_n$, λ_0, moments of Ψ, and structural parameters). Totally differentiating the log of this expression with respect to $\log f_{max}$:

$$\frac{\partial \log g}{\partial \log f_{max}} = \mu \frac{\partial \log M}{\partial \log f_{max}} - \theta\mu. \tag{35}$$

Totally differentiating (34):

$$\frac{\partial \log M}{\partial \log f_{max}} = (1-\varepsilon_D)\frac{\partial \log g}{\partial \log f_{max}}, \tag{36}$$

where $\varepsilon_D \equiv \frac{gD'(g)}{D(g)} = \frac{g[(\phi-1)\log(\gamma)+1]}{D(g)} \in (0,1)$.

Substituting (36) into (35) and solving:

$$\frac{\partial \log g}{\partial \log f_{max}} = \frac{-\theta\mu}{1-\mu(1-\varepsilon_D)}, \qquad \frac{\partial \log M}{\partial \log f_{max}} = \frac{-(1-\varepsilon_D)\theta\mu}{1-\mu(1-\varepsilon_D)}.$$

The denominator $1-\mu(1-\varepsilon_D) > 0$ since $\mu < 1$ (which follows from the equilibrium existence condition $\alpha(\theta+\beta) < 1-\eta$, as this implies $(1-\eta)+\alpha(1-\beta) < 1+\theta$) and $0 < \varepsilon_D < 1$. Therefore $\partial \log M/\partial \log f_{max} < 0$: a decline in expansion costs increases the equilibrium mass of innovative firms.

A.6 Proof of Proposition 6.

A.6.1 Restricted problem

For ease of exposition, we consider first the problem of a restricted planner that takes the mass of innovative firms in the economy as given. In addition, recall that the labor endowments are fixed. Thus, the tool available to the planner is to determine the allocation of resources within the innovation sector between R&D and firm expansion. In such a scenario, the planner wishes to maximize the aggregate arrival rate of ideas, since it raises growth without reallocating resources away from production. The planner's problem is therefore

$$\begin{aligned}
&\max_{\{\ell_n(z),\chi_n(z)\}_{n=1}^{N}} M\int_z \sum_{n=1}^{N} \chi_n(z)\, K_n z \ell_n^{\eta}(z)\, d\Psi(z) \\
&\text{subject to:}\quad K_n = \bar{K}_n \left(M \int_z \chi_n(z)\, z\ell_n(z)^{\beta}\, d\Psi(z) \right)^{\alpha} \\
&\qquad \bar{H}_n = M\int_z \chi_n(z)\left(\bar{f}_n(z) + \ell_n(z)\right) d\Psi(z) \\
&\qquad \bar{f}_n(z) = \frac{\theta}{\theta+1} f_{max} \left(\chi_n(z)\right)^{\frac{1}{\theta}}
\end{aligned}$$

Let $x_n(z) \equiv \chi_n(z)\,\bar{f}_n(z)$ and $y_n(z) \equiv \chi_n(z)\,\ell_n(z)$ denote the amount of region-n innovation labor allocated to expansion and R&D activities of type-z firms, respectively. Let $\psi(z)$ be the PDF of firms of type z in the economy, and define $H_n(z) = \psi(z)(x_n(z) + y_n(z))$ and $\tilde{H}_n(z) = H_n(z)/\psi(z)$. Note that the probability that a firm of type z opens a plant in location n, $\chi_n(z)$, and the expected amount of labor that a firm of type z hires in location n to pay the fixed costs conditional on being active there, $\bar{f}_n(z)$, are linked by the following equation:

$$\chi_n(z) = \left(f_{max} \frac{\theta}{\theta+1} \right)^{-\theta} \bar{f}_n(z)^{\theta}.$$

Combining this expression with the definitions of $x_n(z)$ and $y_n(z)$ we obtain:

$$\bar{f}_n(z) = \left(\frac{\theta}{\theta+1} f_{max} \right)^{\frac{\theta}{\theta+1}} x_n(z)^{\frac{1}{\theta+1}} \tag{37}$$

$$\chi_n(z) = \left(\frac{\theta}{\theta+1} f_{max} \right)^{-\frac{\theta}{\theta+1}} x_n(z)^{\frac{\theta}{\theta+1}} \tag{38}$$

$$y_n(z) = \frac{H_n(z)}{\psi(z)} - x_n(z). \tag{39}$$

Using these definitions, we note that, for any constants $a_1, a_2 > 0$, we obtain:

$$\chi_n(z) z^{a_1} \ell_n(z)^{a_2} = \chi_n(z) z^{a_1} \left(\frac{y_n(z)}{\chi_n(z)} \right)^{a_2} = \left(\frac{\theta}{\theta+1} f_{max} \right)^{-\frac{\theta}{\theta+1}(1-a_2)} x_n(z)^{\frac{\theta}{\theta+1}(1-a_2)} z^{a_1} y_n(z)^{a_2}.$$

We then define the following objects:

$$\begin{aligned}
D_n(z) &\equiv \psi(z) z x_n(z)^{\frac{\theta}{\theta+1}(1-\eta)} y_n(z)^{\eta} \\
B_n(z) &\equiv \psi(z) z x_n(z)^{\frac{\theta}{\theta+1}(1-\beta)} y_n(z)^{\beta}.
\end{aligned}$$

Combining the expressions above, we define the following variables:

$$s_n^{\eta}(z) \equiv \frac{\psi(z)\,\chi_n(z)\,z(\ell_n(z))^{\eta}}{\int_{z'}\psi(z')\,\chi_n(z')\,z'(\ell_n(z'))^{\eta}dz'} = \frac{D_n(z)}{\int_{z'}D_n(z')dz'}$$
$$s_n^{\beta}(z) \equiv \frac{\psi(z)\,\chi_n(z)\,z(\ell_n(z))^{\beta}}{\int_{z'}\psi(z')\,\chi_n(z')\,z'(\ell_n(z'))^{\beta}dz'} = \frac{B_n(z)}{\int_{z'}B_n(z')dz'}.$$

Using these expressions, we can write the following maximization problem, which yields a solution equivalent to the planner's allocation:

$$\max_{x_n(z),y_n(z)} M^{1+\alpha}\sum_{n=1}^{N}\bar{K}_n\left(\int_{z'}B_n(z')dz'\right)^{\alpha}\int_z D_n(z)dz + \\ +\sum_{n=1}^{N}\mathcal{L}_n\left(\bar{H}_n - M\int_z \psi(z)(x_n(z)+y_n(z))dz\right).$$

The first order conditions of this problem with respect to x_n and y_n are:

$$M^{1+\alpha}\bar{K}_n\left[\left(\int_{z'}B_n(z')dz'\right)^{\alpha}\frac{\partial D_n(z)}{\partial x_n(z)} + \alpha\left(\int_{z'}B_n(z')dz'\right)^{\alpha-1}\frac{\partial B_n(z)}{\partial x_n(z)}\int_z D_n(z)dz\right] = \mathcal{L}_n\psi(z)$$
$$M^{1+\alpha}\bar{K}_n\left[\left(\int_{z'}B_n(z')dz'\right)^{\alpha}\frac{\partial D_n(z)}{\partial y_n(z)} + \alpha\left(\int_{z'}B_n(z')dz'\right)^{\alpha-1}\frac{\partial B_n(z)}{\partial y_n(z)}\int_z D_n(z)dz\right] = \mathcal{L}_n\psi(z).$$

Note that

$$\frac{\partial D_n(z)}{\partial x_n(z)} = \frac{\theta}{\theta+1}(1-\eta)\frac{D_n(z)}{x_n(z)} \quad \text{and} \quad \frac{\partial D_n(z)}{\partial y_n(z)} = \eta\frac{D_n(z)}{y_n(z)}$$
$$\frac{\partial B_n(z)}{\partial x_n(z)} = \frac{\theta}{\theta+1}(1-\beta)\frac{B_n(z)}{x_n(z)} \quad \text{and} \quad \frac{\partial B_n(z)}{\partial y_n(z)} = \beta\frac{B_n(z)}{y_n(z)}$$

Thus, the first order conditions can be rewritten as:

$$M^{1+\alpha}\frac{\theta}{\theta+1}\frac{\bar{K}_n\left(\int_{z'}B_n(z')dz'\right)^{\alpha}\int_z D_n(z)dz}{x_n(z)}\left[(1-\eta)s_n^{\eta}(z) + \alpha(1-\beta)s_n^{\beta}(z)\right] = \mathcal{L}_n\psi(z) \tag{40}$$
$$M^{1+\alpha}\frac{\bar{K}_n\left(\int_{z'}B_n(z')dz'\right)^{\alpha}\int_z D_n(z)dz}{y_n(z)}\left[\eta s_n^{\eta}(z) + \alpha\beta s_n^{\beta}(z)\right] = \mathcal{L}_n\psi(z) \tag{41}$$

Taking the ratio of these two equations we find that:

$$\frac{x_n(z)}{y_n(z)} = \frac{\theta}{\theta+1}\frac{\left[(1-\eta)s_n^{\eta}(z) + \alpha(1-\beta)s_n^{\beta}(z)\right]}{\left[\eta s_n^{\eta}(z) + \alpha\beta s_n^{\beta}(z)\right]}$$
$$= \frac{\left(\frac{s_n^{\beta}(z)}{s_n^{\eta}(z)}\right)\alpha\frac{1-\beta}{1-\eta}+1}{\left(\frac{s_n^{\beta}(z)}{s_n^{\eta}(z)}\right)\alpha\frac{\beta}{\eta}+1}\left(\frac{1-\eta}{\eta}\frac{\theta}{\theta+1}\right). \tag{42}$$

A.6.2 Generalized welfare function

Next we turn to the general case where the planner can choose the number of innovative firms M. The planner takes into account the fixed cost of opening an innovative firm, which is $f_e Y_t$ in terms of the final good. Increasing innovative firms' entry thus reduces aggregate consumption, C_t, which is equal to final good output minus the resources spent on fixed costs: $C_t = Y_t - M f_e Y_t$.

Consider the unrestricted problem of a planner along a balanced growth path. The planner maximizes the following welfare function, which aggregates the utility of each household in location n and sector s at time t with some non-negative weights $\{\{\nu_{nt}^s\}_{s\in\{p,i\}}\}_{n=1}^N$:

$$W_0 = \int_0^\infty e^{-\rho t} \sum_{n=1}^{N} \sum_{s=\{p,i\}} \nu_{nt}^s u\left(C_{ns0}e^{gt}\right) dt,$$

where $C_{nst} = C_{ns0}e^{gt}$ is the amount of final-good consumption in period t for each household, summing up to total aggregate consumption in each period: $C_t = \sum_{n=1}^N \sum_{s\in\{p,i\}} C_{nst}$. Note that for a given amount of final good dedicated to consumption, C_t, the planner could choose transfers to redistribute resources across households to maximize welfare. We are not interested in solving for those transfers, but in understanding the optimal allocation of resources between R&D and expansion. As a consequence, note that we can equivalently express aggregate welfare as a function of aggregate consumption:

$$W_0 = \int_0^\infty e^{-\rho t} v\left(C_t\right) dt,$$

where the function $v(\cdot)$ is strictly increasing and concave as long as $u(\cdot)$ is strictly increasing and concave and the weights $\{\{\nu_{nt}^s\}_{s\in\{p,i\}}\}_{n=1}^N$ are non-negative.

As a result, using the expression from equation (31), the problem of the planner becomes:

$$\max_{\{\ell_n(z),\chi_n(z)\}_{n=1}^N, M} \int_0^\infty e^{-\rho t} v\left(\mathcal{A}_0 \mathcal{L}^p (1 - M f_e) e^{gt}\right) dt$$

$$\text{s.t.} \quad g = \log(\gamma) M \int_z \sum_{n=1}^N \chi_n\left(z\right) K_n z \ell_n^\eta\left(z\right) d\Psi\left(z\right)$$

$$K_n = \bar{K}_n \left(M \int_z \chi_n\left(z\right) z \ell_n\left(z\right)^\beta d\Psi\left(z\right)\right)^\alpha$$

$$\bar{H}_n = M \int_z \chi_n\left(z\right) \left(\bar{f}_n\left(z\right) + \ell_n\left(z\right)\right) d\Psi\left(z\right)$$

$$\bar{f}_n\left(z\right) = \frac{\theta}{\theta+1} f_{max}\left(\chi_n\left(z\right)\right)^{\frac{1}{\theta}}.$$

Using the expressions for $B_n(z)$ and $D_n(z)$ introduced in the previous section, we can write

$$K_n = \left(\frac{\theta}{\theta+1} f_{max}\right)^{-\frac{\theta}{\theta+1}\alpha(1-\beta)} \bar{K}_n M^{\alpha} \left(\int_{z'} B_n(z')dz'\right)^{\alpha}$$

and we can express the growth rate as

$$g = \left(\frac{\theta}{\theta+1} f_{max}\right)^{-\frac{\theta}{\theta+1}(\alpha(1-\beta)+1-\eta)} \log(\gamma) M^{1+\alpha} \sum_{n=1}^{N} \bar{K}_n \left(\int_{z'} B_n(z')dz'\right)^{\alpha} \int_z D_n(z)dz.$$

Using these expressions, we can write the following maximization problem, which yields a solution equivalent to the planner's allocation:

$$\max_{x_n(z), y_n(z), M} \int_0^{\infty} e^{-\rho t} v\left(\mathcal{A}_0 \mathcal{L}^p (1 - M f_e) e^{gt}\right) dt +$$
$$+ \mu \left(g - \left(\frac{\theta}{\theta+1} f_{max}\right)^{-\frac{\theta}{\theta+1}(\alpha(1-\beta)+1-\eta)} \log(\gamma) M^{1+\alpha} \sum_{n=1}^{N} \bar{K}_n \left(\int_{z'} B_n(z')dz'\right)^{\alpha} \int_z D_n(z)dz\right)$$
$$+ \sum_{n=1}^{N} \mu_n \left(H_n - M \int_z \psi(z)(x_n(z) + y_n(z))dz\right).$$

The first order conditions of this problem are:

$$-\mu \frac{\partial g}{\partial x_n(z)} - \mu_n M \psi(z) = 0 \tag{43}$$

$$-\mu \frac{\partial g}{\partial y_n(z)} - \mu_n M \psi(z) = 0 \tag{44}$$

$$-\int_0^{\infty} e^{(g-\rho)t} v'(C_t) \left(\mathcal{A}_0 \mathcal{L}^p f_e\right) dt - \mu(1+\alpha)\frac{g}{M} - \mu_n \left(\int_z \psi(z)(x_n(z) + y_n(z))dz\right) = 0 \tag{45}$$

Note that equations (43) and (44) have the same solution as the first order conditions in the restricted planner problem in equations (40) and (41). As a result, the relative allocation of resources between R&D and expansion, for a given level of firm entry M, is the same in the restricted and unrestricted problem, and it satisfies Equation (42).

A.7 Proof of Proposition 7.

To derive the planner's allocation across firm types, we continue with the analysis presented in Section A.6.1. Solving the first order conditions in equations (40) and (41) for $x_n(z)$ and $y_n(z)$ we can solve for the mass of labor hired by firms of type z in market n, $H_n(z)$:

$$\begin{aligned}H_n(z) &= M\psi(z)(x_n(z)+y_n(z)) = \\ &= M^{1+\alpha}\frac{\bar{K}_n\left(\int_{z'}B_n(z')dz'\right)^{\alpha}\int_z D_n(z)dz}{\mathcal{L}_n}\times \\ &\qquad\times\left(\frac{\theta}{\theta+1}\left[(1-\eta)s_n^{\eta}(z)+\alpha(1-\beta)s_n^{\beta}(z)\right]+\left[\eta s_n^{\eta}(z)+\alpha\beta s_n^{\beta}(z)\right]\right)\end{aligned}$$

As a result, the share of local R&D labor allocated to type-z firms in market n is:

$$\begin{aligned}\frac{H_n(z)}{H_n} &= \frac{\frac{\theta}{\theta+1}\left[(1-\eta)s_n^{\eta}(z)+\alpha(1-\beta)s_n^{\beta}(z)\right]+\left[\eta s_n^{\eta}(z)+\alpha\beta s_n^{\beta}(z)\right]}{\int_{z'}\left(\frac{\theta}{\theta+1}\left[(1-\eta)s_n^{\eta}(z)+\alpha(1-\beta)s_n^{\beta}(z)\right]+\left[\eta s_n^{\eta}(z)+\alpha\beta s_n^{\beta}(z)\right]\right)d\Psi(z')} \\ &= \frac{\alpha s_n^{\beta}(z)\left(\beta+(1-\beta)\frac{\theta}{\theta+1}\right)+s_n^{\eta}(z)\left(\eta+(1-\eta)\frac{\theta}{\theta+1}\right)}{\int_{z'}\left(\alpha s_n^{\beta}(z')\left(\beta+(1-\beta)\frac{\theta}{\theta+1}\right)+s_n^{\eta}(z')\left(\eta+(1-\eta)\frac{\theta}{\theta+1}\right)\right)d(z')}.\end{aligned}$$

A.8 Proof of Proposition 8.

In this section, we examine the planner's allocation under free labor mobility in the innovation sector. The planner's problem is

$$\begin{aligned}&\max_{\{H_n,\ell_n(z),\chi_n(z)\}_{n=1}^{N}} M\int_z\sum_{n=1}^{N}\chi_n(z)K_n z\ell_n^{\eta}(z)d\Psi(z) \\ &\text{subject to:}\quad K_n=\bar{K}_n\left(M\int_z\chi_n(z)z\ell_n(z)^{\beta}d\Psi(z)\right)^{\alpha} \\ &\qquad H_n=M\int_z\chi_n(z)\left(\bar{f}_n(z)+\ell_n(z)\right)d\Psi(z) \\ &\qquad \bar{f}_n(z)=\frac{\theta}{\theta+1}f_{max}(\chi_n(z))^{\frac{1}{\theta}} \\ &\qquad \sum_{n=1}^{N}H_n=\bar{H}.\end{aligned}$$

To solve the planner's problem, we define $x_n(z)\equiv\chi_n(z)\bar{f}_n(z)$ and $y_n(z)\equiv\chi_n(z)\ell_n(z)$ as the amount of region-n innovation labor allocated to expansion and R&D activities of type-z firms, respectively, as in Section A.6. We also use the definitions of $D_n(z)$, $B_n(z)$, $s_n^{\eta}(z)$, and $s_n^{\beta}(z)$ introduced in Section A.6.

Using these expressions, we can write the following maximization problem, which yields a solution equivalent to the planner's allocation:

$$\max_{x_n(z), y_n(z), H_n} M^{1+\alpha} \sum_{n=1}^{N} \bar{K}_n \left(\int_{z'} B_n(z')dz' \right)^{\alpha} \int_z D_n(z)dz +$$
$$+ \sum_{n=1}^{N} \mathcal{L}_n \left(H_n - M \int_z \psi(z)(x_n(z) + y_n(z))dz \right) +$$
$$+ \bar{\mathcal{L}} \left(\bar{H} - \sum_{n=1}^{N} H_n \right).$$

The first order conditions of this problem with respect to x_n, y_n, and H_n are:

$$M^{1+\alpha} \bar{K}_n \left[\left(\int_{z'} B_n(z')dz' \right)^{\alpha} \frac{\partial D_n(z)}{\partial x_n(z)} + \alpha \left(\int_{z'} B_n(z')dz' \right)^{\alpha-1} \frac{\partial B_n(z)}{\partial x_n(z)} \int_z D_n(z)dz \right] = \mathcal{L}_n \psi(z)$$
$$M^{1+\alpha} \bar{K}_n \left[\left(\int_{z'} B_n(z')dz' \right)^{\alpha} \frac{\partial D_n(z)}{\partial y_n(z)} + \alpha \left(\int_{z'} B_n(z')dz' \right)^{\alpha-1} \frac{\partial B_n(z)}{\partial y_n(z)} \int_z D_n(z)dz \right] = \mathcal{L}_n \psi(z)$$
$$\mathcal{L}_n = \bar{\mathcal{L}}.$$

Note that

$$\frac{\partial D_n(z)}{\partial x_n(z)} = \frac{\theta}{\theta+1}(1-\eta)\frac{D_n(z)}{x_n(z)} \qquad \text{and} \qquad \frac{\partial D_n(z)}{\partial y_n(z)} = \eta \frac{D_n(z)}{y_n(z)}$$
$$\frac{\partial B_n(z)}{\partial x_n(z)} = \frac{\theta}{\theta+1}(1-\beta)\frac{B_n(z)}{x_n(z)} \qquad \text{and} \qquad \frac{\partial B_n(z)}{\partial y_n(z)} = \beta \frac{B_n(z)}{y_n(z)}$$

Thus equation the first order conditions can be rewritten as:

$$M^{1+\alpha} \frac{\theta}{\theta+1} \frac{\bar{K}_n \left(\int_{z'} B_n(z')dz' \right)^{\alpha} \int_z D_n(z)dz}{x_n(z)} \left[(1-\eta)s_n^{\eta}(z) + \alpha(1-\beta)s_n^{\beta}(z) \right] = \mathcal{L}_n \psi(z)$$
$$M^{1+\alpha} \frac{\bar{K}_n \left(\int_{z'} B_n(z')dz' \right)^{\alpha} \int_z D_n(z)dz}{y_n(z)} \left[\eta s_n^{\eta}(z) + \alpha\beta s_n^{\beta}(z) \right] = \mathcal{L}_n \psi(z)$$

Taking the ratio of these two equations we find that:

$$\frac{x_n(z)}{y_n(z)} = \frac{\theta}{\theta+1} \frac{\left[(1-\eta)s_n^{\eta}(z) + \alpha(1-\beta)s_n^{\beta}(z) \right]}{\left[\eta s_n^{\eta}(z) + \alpha\beta s_n^{\beta}(z) \right]}.$$

Solving these equations for $x_n(z)$ and $y_n(z)$ and plugging them into the local labor market clearing equation, we get:

$$H_n = M \int_z \psi(z)(x_n(z) + y_n(z))dz =$$
$$= M^{1+\alpha} \int_z \frac{\bar{K}_n \left(\int_{z'} B_n(z')dz'\right)^{\alpha} \int_z D_n(z)dz}{\mathcal{L}_n} \times$$
$$\times \left(\frac{\theta}{\theta+1} \left[(1-\eta)s_n^{\eta}(z) + \alpha(1-\beta)s_n^{\beta}(z)\right] + \left[\eta s_n^{\eta}(z) + \alpha\beta s_n^{\beta}(z)\right] \right) dz$$

Define the following variable:

$$\Upsilon_n \equiv \bar{K}_n \left(\int_{z'} B_n(z')dz' \right)^{\alpha} \int_z D_n(z)dz.$$

Note that $\int_z s_n^{\eta}(z)dz = \int_z s_n^{\beta}(z)dz = 1$. It follows that:

$$H_n = M^{1+\alpha} \frac{\Upsilon_n}{\bar{\mathcal{L}}} \left(\frac{\theta}{\theta+1} \left((1-\eta) + \alpha(1-\beta)\right) + (\eta + \alpha\beta) \right).$$

Combining the expressions for x_n, y_n, B_n, and D_n, note that

$$\Upsilon_n \propto \bar{K}_n \Upsilon_n^{\alpha\left[\frac{\theta}{\theta+1}(1-\beta)+\beta\right] + \frac{\theta}{\theta+1}(1-\eta)+\eta}.$$

Rearranging this expression, we obtain that:

$$H_n \propto \Upsilon_n \propto \bar{K}_n^{\frac{1}{\frac{1-\eta}{1+\theta} - \alpha\frac{\beta+\theta}{\theta+1}}}.$$

In the decentralized equilibrium, from the model inversion, we obtain that:

$$\bar{K}_n \propto w_{ni} \left(\bar{H}_n\right)^{\frac{1-\eta}{1+\theta} - \alpha\frac{\beta+\theta}{\theta+1}}$$

As a result, under labor mobility, when wages are equalized across locations, the allocations of innovative labor in the decentralized equilibrium and in the planner's solution are the same. In particular, in both the equilibrium and the planner' allocation, we obtain that:

$$\frac{H_n}{\sum_{n'=1}^{N} H_{n'}} = \frac{\bar{K}_n^{\frac{1}{\frac{1-\eta}{1+\theta} - \alpha\frac{\beta+\theta}{\theta+1}}}}{\sum_{n'=1}^{N} \bar{K}_{n'}^{\frac{1}{\frac{1-\eta}{1+\theta} - \alpha\frac{\beta+\theta}{\theta+1}}}}.$$

A.9 Efficiency of production worker spatial allocation

In footnote 16, we note that the spatial allocation of production workers is efficient. To show this result, we include the allocation of production labor as an additional control variable in the planner's problem stated in equation (12).

The planner's problem in equation (12) maximizes aggregate welfare subject to constraints on innovation activities. To analyze the allocation of production workers, we extend this problem to include the allocation of production labor $\{L_{np}\}_{n=1}^N$ across space, subject to an additional resource constraint:

$$\sum_{n=1}^{N} L_{np} = \bar{L} \tag{46}$$

where $\bar{L} = \sum_{n=1}^{N} \bar{L}_n$ is the aggregate endowment of production workers.

Following the approach in Section A.6.2, we note that the planner can redistribute the final good across households. Thus, we can equivalently express aggregate welfare as a function of aggregate consumption:

$$W_0 = \int_0^\infty e^{-\rho t} v(C_t) dt$$

where the function $v(\cdot)$ is strictly increasing and concave.

Using the expression in equation (31), aggregate output is $Y_t = \mathcal{A}_t \mathcal{L}^p$ where $\mathcal{A}_t$ is aggregate productivity and $\mathcal{L}^p$ is the aggregate production labor index defined by:

$$\log(\mathcal{L}^p) \equiv \int_0^1 \log \left[\sum_{n=1}^{N} (Z_n L_{jn})^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} dj.$$

The planner's problem becomes:

$$\max_{\{\ell_n(z), \chi_n(z)\}_{n=1}^N, M, \{L_{np}\}_{n=1}^N} \quad \int_0^\infty e^{-\rho t} v\left(\mathcal{A}_0 \mathcal{L}^p (1 - M f_e) e^{gt}\right) dt$$

subject to the constraints in equation (12), plus constraint (46).

The growth rate g depends only on innovation activities through $\{\chi_n(z), \ell_n(z)\}_{n=1}^N$ and M. The spillover function K_n similarly depends only on innovation variables. Therefore, conditional on the allocation of resources to innovation (which determines g and $\mathcal{A}_t$), the planner's choice of production labor allocation $\{L_{np}\}_{n=1}^N$ affects aggregate welfare only through the production labor index $\mathcal{L}^p$.

Given this separability, we can analyze the planner's optimal allocation of production labor by maximizing $\mathcal{L}^p$ subject to the aggregate labor constraint (46). Since the final good aggregator is Cobb-Douglas (equation 1) and all sectors share the same CES production function over locations (differing only in their productivity level A_j), the planner's optimal spatial allocation L_{jn}/L_j is identical across sectors j. Intuitively, A_j scales sector j's output but does not affect the optimal mix of labor across locations. As a result, we can write $L_{jn} = (L_j/\bar{L})\, L_{np}$ for all j, and the planner's sub-problem reduces to:

$$\max_{\{L_{np}\}_{n=1}^{N}} \quad \left[\sum_{n=1}^{N}(Z_n L_{np})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \qquad \text{s.t.} \quad \sum_{n=1}^{N} L_{np} = \bar{L}.$$

The first-order condition with respect to L_{np} requires:

$$Z_n^{\frac{\sigma-1}{\sigma}} L_{np}^{-\frac{1}{\sigma}} = \xi \quad \forall n$$

where ξ is the Lagrange multiplier on constraint (46). This states that the planner equalizes the marginal product of production labor across all locations. Taking ratios across any two locations n and n':

$$\frac{Z_n^{\frac{\sigma-1}{\sigma}} L_{np}^{-\frac{1}{\sigma}}}{Z_{n'}^{\frac{\sigma-1}{\sigma}} L_{n'p}^{-\frac{1}{\sigma}}} = 1 \quad \forall n, n'. \tag{47}$$

In the decentralized equilibrium, intermediate goods producers choose production labor across locations to minimize costs. From equation (2), the cost-minimization problem for intermediate good j is:

$$\min_{\{L_{jn}\}_{n=1}^{N}} \quad \sum_{n=1}^{N} w_{np} L_{jn} \qquad \text{s.t.} \quad y_j = A_j \left[\sum_{n=1}^{N}(Z_n L_{jn})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}.$$

The first-order condition with respect to L_{jn} yields:

$$w_{np} = \lambda_j A_j \left[\sum_{n'=1}^{N}(Z_{n'} L_{jn'})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} Z_n^{\frac{\sigma-1}{\sigma}} L_{jn}^{-\frac{1}{\sigma}}$$

where λ_j is the Lagrange multiplier on the production constraint. Taking ratios across locations n and n' (which cancels λ_j, A_j, and the CES aggregate):

$$\frac{w_{np}}{w_{n'p}} = \frac{Z_n^{\frac{\sigma-1}{\sigma}} L_{jn}^{-\frac{1}{\sigma}}}{Z_{n'}^{\frac{\sigma-1}{\sigma}} L_{jn'}^{-\frac{1}{\sigma}}}.$$

Since the left-hand side is independent of j, all sectors allocate labor across locations in the same proportions, so $L_{jn}/L_j = L_{np}/\bar{L}$ for all j, and therefore:

$$\frac{w_{np}}{w_{n'p}} = \frac{Z_n^{\frac{\sigma-1}{\sigma}} L_{np}^{-\frac{1}{\sigma}}}{Z_{n'}^{\frac{\sigma-1}{\sigma}} L_{n'p}^{-\frac{1}{\sigma}}}. \tag{48}$$

Under free mobility of production workers, wages equalize across locations: $w_{np} = w_{n'p}$ for all

n, n'. Equation (48) then implies:

$$\frac{Z_n^{\frac{\sigma-1}{\sigma}} L_{np}^{-\frac{1}{\sigma}}}{Z_{n'}^{\frac{\sigma-1}{\sigma}} L_{n'p}^{-\frac{1}{\sigma}}} = 1 \quad \forall n, n',$$

which is identical to the planner's optimality condition in equation (47). Since the CES aggregator is strictly concave, this condition pins down a unique allocation of $\{L_{np}\}_{n=1}^{N}$. Therefore, the decentralized equilibrium allocation of production labor across space coincides with the planner's optimal allocation. The competitive labor market achieves efficiency because firms fully internalize the marginal product of production workers, and there are no externalities in production.

A.10 Additional details for Section 2.5

A.10.1 Spillovers from all regions

The model outlined so far assumed that spillovers operate only at the local level, i.e. innovative firms gain insights only from firms that share their locations. This framework can easily be extended to allow for spillovers from all regions, where frictions for knowledge to travel across space from location i to location n are captured by a parameter κ_{in}. Let the local knowledge spillover K_n now be given by

$$K_n = \bar{K}_n \left(\sum_{i=1}^{N} \int_{\omega \in \Omega_i} \kappa_{in} z_\omega \ell_{i\omega}^{\beta} d\omega \right)^{\alpha}. \tag{49}$$

Under this assumption, the optimal allocation is the same, with

$$s_n^{\beta}(z) \equiv \sum_{k=1}^{N} \int_{z'} s_k^{\eta}(z')\, s_{nk}^{\beta}(z)\, d\Psi(z') \tag{50}$$

where $s_{nk}^{\beta}(z)$ is the weight of plants from region n, type z in delivering insights to k.

A.10.2 Span of control

In our model, innovative firms need to pay an idiosyncratic fixed cost to open an establishment in a given location. An alternative assumption commonly used in the spatial literature is that, instead of random fixed costs, geographic expansion is costly due to a span of control problem. In this section, we explore the properties of our model under this alternative assumption.

In particular, suppose that space is continuous and symmetric, and that innovative firms are homogeneous, so we drop the indexes n and ω and the firm-specific R&D productivity z_ω. We focus on a balanced growth path and drop the time index. We consider the case where innovative labor is perfectly mobile, so there is a unique wage w_i. Instead of random fixed costs, an innovative firm operating in $\mathcal{N}$ locations incurs a cost of expansion given by $(\mathcal{N})^{1+\frac{1}{\theta}} w_i$. Under these assumptions, we show that the optimal allocation between R&D and expansion follows the same expression as in Proposition 6.

The innovative firm problem takes the following form:

$$\max_{\mathcal{N},\ell} \lambda V - \mathcal{N} w_i \ell - w_i \mathcal{N}^{1+\frac{1}{\theta}} \quad \text{s.t.} \quad \lambda = \lambda_0 \left(\mathcal{N} K \ell^{\eta}\right)$$

The first order conditions of the firm's problem are:

$$\lambda_0 K \ell^{\eta} V = w_i \ell + \left(1 + \frac{1}{\theta}\right) w_i \mathcal{N}^{\frac{1}{\theta}}$$
$$\eta \lambda_0 \mathcal{N} K \ell^{\eta-1} V = w_i \mathcal{N}$$

Rearranging these equations, we obtain that the decentralized equilibrium allocation of innovative resources between expansion and R&D satisfies:

$$\frac{\mathcal{N}^{1+\frac{1}{\theta}}}{\ell \mathcal{N}} = \frac{1-\eta}{\eta} \frac{\theta}{\theta+1}.$$

Note that the allocation of innovative resources is the same as in our baseline model with random fixed costs, as described in Proposition 2.

We next consider the restricted problem of a planner that takes the mass of innovative firms as given as in Section A.6.1. Under the assumptions outlined above, the planner's problem takes the following form:

$$\max_{\mathcal{N}_p,\ell_p} M\bar{\lambda}$$
$$\text{s.t.} \bar{\lambda} = \lambda_0 \left(\mathcal{N}_p K \ell_p^{\eta}\right)$$
$$K = \bar{K} M^{\alpha} \mathcal{N}_p^{\alpha} \ell_p^{\beta\alpha}$$
$$M(\mathcal{N}_p^{\delta} + \mathcal{N}_p \ell_p) = \bar{H}$$

Letting $\mathcal{L}$ be the Lagrange multiplier on the innovative labor resource constraint, the first order conditions of the planner's problem are:

$$(\alpha+1)\bar{K} M^{1+\alpha} \lambda_0 \mathcal{N}_p^{\alpha} \ell_p^{\alpha\beta+\eta} = \mathcal{L} M \left(\left(1+\frac{1}{\theta}\right) \mathcal{N}_p^{\frac{1}{\theta}} + \ell_p\right)$$
$$(\alpha\beta+\eta)\bar{K} M^{1+\alpha} \lambda_0 \mathcal{N}_p^{\alpha+1} \ell_p^{\alpha\beta+\eta-1} = \mathcal{L} M \mathcal{N}_p$$

Rearranging these equations, we find that the planner's allocation of innovative resources between expansion and R&D satisfies:

$$\frac{\mathcal{N}_p^{1+\frac{1}{\theta}}}{\ell_p \mathcal{N}_p} = \frac{\theta}{\theta+1} \frac{1-\eta+\alpha(1-\beta)}{\eta+\alpha\beta}.$$

Note that this is the same expression we obtain in the baseline model with idiosyncratic fixed costs, as outlined in Proposition 6. As in our baseline model, it follows that the planner's allocation is the same as the decentralized equilibrium when $\alpha = 0$, which is the case when there are no knowledge spillovers. If $\alpha > 0$, the planner allocates more resources to expansion relative to the competitive equilibrium when $\eta > \beta$.

A.10.3 Imperfect substitution across locations

In our baseline model, multi-location innovative firms aggregate the innovation output of each of their locations linearly, as described in equation (3). In this section, we consider an alternative specification where different locations are imperfect substitutes in the firm's innovation production function:

$$\lambda_\omega = \lambda_0 \left[\sum_{n=1}^{N} K_n \left(z_\omega \ell_{n\omega}^{\eta} \right)^{\frac{\zeta-1}{\zeta}} \right]^{\frac{\zeta}{\zeta-1}}.$$

For simplicity, let us assume that space is continuous and symmetric, as in Section A.10.2. Then the problem of the firm becomes:

$$\max_{\{\ell_\omega, N_\omega\}} V\lambda_0 \left(N_\omega K\right)^{\frac{\zeta}{\zeta-1}} \left(z_\omega \ell_\omega^{\eta}\right) - w_i N_\omega \left(\frac{\theta}{\theta+1} f_{max} \left(N_\omega\right)^{\frac{1}{\theta}} + \ell_\omega \right)$$

The first order conditions of this problem are:

$$\eta V\lambda_0 \left(N_\omega K\right)^{\frac{\zeta}{\zeta-1}} \left(z_\omega \ell_\omega^{\eta-1}\right) = w_i N_\omega$$

$$\frac{\zeta}{\zeta-1} V\lambda_0 \left(N_\omega\right)^{\frac{\zeta}{\zeta-1}-1} \left(K\right)^{\frac{\zeta}{\zeta-1}} \left(z_\omega \ell_\omega^{\eta}\right) = w_i \left(\frac{\theta+1}{\theta} N_\omega^{\frac{1}{\theta}} \frac{\theta}{\theta+1} f_{max} + \ell_\omega \right)$$

Rearranging these equations we obtain that the allocation of innovative labor to expansion relative to R&D in equilibrium satisfies the following equation:

$$\frac{N_\omega \frac{\theta}{\theta+1} f_{max} N_\omega^{\frac{1}{\theta}}}{N_\omega \ell_\omega} = \frac{\theta}{\theta+1} \frac{1 - \frac{\zeta-1}{\zeta}\eta}{\frac{\zeta-1}{\zeta}\eta}. \tag{51}$$

Now, in the same setting, we consider the problem of the planner, who internalizes the spillovers in the composition of R&D efficiency in each location:

$$\max_{\{\ell_\omega, N_\omega\}} V\lambda_0 \left(N_\omega \bar{K} \left(M N_\omega z_\omega \ell_\omega^{\beta} \right)^{\alpha} \right)^{\frac{\zeta}{\zeta-1}} \left(z_\omega \ell_\omega^{\eta}\right) - w_i N_\omega \left(\frac{\theta}{\theta+1} f_{max} \left(N_\omega\right)^{\frac{1}{\theta}} + \ell_\omega \right)$$

The first order conditions of this problem are:

$$\left(\alpha\beta\frac{\zeta}{\zeta-1}+\eta\right)V\lambda_0\left(N_\omega\right)^{(1+\alpha)\frac{\zeta}{\zeta-1}}\left(\bar{K}M^\alpha\right)^{\frac{\zeta}{\zeta-1}}\left(z_\omega\right)^{\alpha\frac{\zeta}{\zeta-1}+1}\left(\ell_\omega^{\alpha\beta\frac{\zeta}{\zeta-1}+\eta-1}\right)=w_iN_\omega$$
$$\left((1+\alpha)\frac{\zeta}{\zeta-1}\right)V\lambda_0\left(N_\omega\right)^{(1+\alpha)\frac{\zeta}{\zeta-1}-1}\left(\bar{K}M^\alpha\right)^{\frac{\zeta}{\zeta-1}}\left(z_\omega\right)^{\alpha\frac{\zeta}{\zeta-1}+1}\left(\ell_\omega^{\alpha\beta\frac{\zeta}{\zeta-1}+\eta}\right)=w_i\left(\frac{\theta+1}{\theta}N_\omega^{\frac{1}{\theta}}\frac{\theta}{\theta+1}f_{max}+\ell_\omega\right)$$

Rearranging these equations, we find the planner's allocation of innovative labor to expansion relative to R&D stasifies the following expression:

$$\frac{N_\omega\frac{\theta}{\theta+1}f_{max}N_\omega^{\frac{1}{\theta}}}{N_\omega\ell_\omega}=\frac{\theta}{\theta+1}\frac{\alpha(1-\beta)+1-\frac{\zeta-1}{\zeta}\eta}{\alpha\beta+\frac{\zeta-1}{\zeta}\eta}. \tag{52}$$

Equations (51) and (52) illustrate how the planner's relative expenditure on fixed plant costs diverges from that in equilibrium. Note that we recover a result similar to the one in the baseline model in the main text, presented in Proposition 6. In particular, the equilibrium allocation coincides with the planner's when there are no knowledge spillovers ($\alpha=0$). In addition, when β is sufficiently low relative to η, firms in equilibrium under-invest in expansion compared to the planner's optimal solution. Specifically, in this version of the model this is true when $\beta<\eta\frac{\zeta-1}{\zeta}$. Conversely, as in the main model, when β is high relatively to η, firms over-invest in expansion.

A.10.4 General dependence of spillovers on firm productivity

In our baseline model, firm productivity z_ω enters both the innovation production function and the spillover function with the same exponent (equal to one). In this section, we consider a generalization in which the spillover contribution of a firm depends on its productivity raised to a general power $\xi\geq 0$, while the firm's own innovation output retains its baseline dependence on z_ω. Specifically, the knowledge spillover function becomes

$$K_{nt}=\bar{K}_n\left(\int_{\omega\in\Omega_n}z_\omega^\xi\ell_{n\omega t}^\beta\,d\omega\right)^\alpha,$$

while the firm's innovation production function remains $\lambda_{\omega t}=\lambda_0\sum_{n=1}^N K_{nt}z_\omega\ell_{n\omega t}^\eta$. When $\xi=1$, we recover the baseline model. When $\xi<1$, more productive firms contribute relatively less to spillovers per unit of productivity; when $\xi>1$, they contribute relatively more.

Equilibrium. Since the firm takes K_{nt} as given when choosing its labor allocation, the parameter ξ does not affect the firm's private optimization problem. Therefore, the equilibrium allocation of innovation labor between expansion and R&D is identical to the baseline:

$$\frac{x_{nt}(z)}{y_{nt}(z)}=\frac{1-\eta}{\eta}\frac{\theta}{\theta+1}.$$

Planner's allocation. We solve the full planner's problem with heterogeneous firms and locations, following the same approach as in the proof of Proposition 6. The planner's problem is identical to (12), with the modified spillover function above. Define the generalized spillover share:

$$s_n^{\xi,\beta}(z) \equiv \frac{\psi(z)\, z^{\xi}\, x_n(z)^{\frac{\theta}{\theta+1}(1-\beta)}\, y_n(z)^{\beta}}{\int_{z'} \psi(z')\,(z')^{\xi}\, x_n(z')^{\frac{\theta}{\theta+1}(1-\beta)}\, y_n(z')^{\beta}\, dz'},$$

which replaces $s_n^{\beta}(z)$ from Proposition 6 and uses z^{ξ} instead of z in the spillover contribution. The innovation share $s_n^{\eta}(z)$ is unchanged.

Taking first-order conditions with respect to $\ell_n(z)$ and $\chi_n(z)$ and forming their ratio (following the same steps as in the proof of Proposition 6), we obtain the planner's allocation:

$$\frac{x_n(z)}{y_n(z)} = \frac{\alpha \frac{s_n^{\xi,\beta}(z)}{s_n^{\eta}(z)} \frac{1-\beta}{1-\eta} + 1}{\alpha \frac{s_n^{\xi,\beta}(z)}{s_n^{\eta}(z)} \frac{\beta}{\eta} + 1} \frac{1-\eta}{\eta} \frac{\theta}{\theta+1}.$$

This has exactly the same structure as Proposition 6, with $s_n^{\xi,\beta}(z)$ replacing $s_n^{\beta}(z)$. The comparison between the planner's and the equilibrium allocation depends on whether

$$\frac{1-\beta}{1-\eta} \gtrless \frac{\beta}{\eta},$$

which simplifies to $\eta \gtrless \beta$. Crucially, ξ enters only through the shares $s_n^{\xi,\beta}(z)$ but does not affect the direction of the comparison: the planner allocates more resources to expansion than the equilibrium when $\beta < \eta$, less when $\beta > \eta$, and the same when $\beta = \eta$, regardless of ξ.

B Data

Throughout the paper, patents, citations, and inventor locations come from PatentsView. The merged D&B data are used to cross-validate the geographic presence of patenting firms; to construct the inventor-to-firm links used in the patenting-efficiency and estimation exercises; and to provide an alternative employment measure in some estimation specifications.

B.1 Details of merging PatentsView and Dun & Bradstreet

Table B.1: **Example patent record**

Patent ID	Raw Assignee Organization	CZ1990	Grant Date
4310440	Union Carbide Corporation	19400	1982-01-12

Note: This table shows an example patent record.

Table B.4 summarizes the firm-level moments of the merged panel used throughout the paper. These moments complement Table B.3: while Table B.3 shows that the merge preserves broad coverage across patents, assignees, and locations, Table B.4 shows the typical scale and skewness

Table B.2: **Example Dun & Bradstreet record**

dunsno	hqdunsno	parentdunsno	ultdunsno	status	companyname	nationalcode	cz1990	firstyear	lastyear
1289008	1289008	—	—	1	UNION CARBIDE CORP	0	19400	$-\infty$	1970
1289008	1289008	—	1289008	1	UNION CARBIDE CORP	0	19400	1971	1973
1289008	1289008	—	1289008	1	UNION CARBIDE CORPORATION	0	19400	1974	1976
1289008	1289008	—	1289008	1	UNION CARBIDE CORPORATION	0	19400	1978	1981
1289008	1289008	—	1289008	1	UNION CARBIDE CORPORATION	0	20901	1982	1988
1289008	1289008	—	1289008	1	UNION CARBIDE CORPORATION	0	20901	1995	2000
1289008	1289008	1381581	1381581	1	UNION CARBIDE CORPORATION	0	20901	2001	2004
1289008	1289008	1381581	1381581	1	UNION CARBIDE CORPORATION	0	32000	2005	2016
1289008	1289008	1381581	80386615	1	UNION CARBIDE CORPORATION	0	32000	2017	2018
1289008	1289008	1381581	117063457	1	UNION CARBIDE CORPORATION	0	32000	2019	∞
1289008	1289008	602844128	602844128	1	UNION CARBIDE CHEM & PLAS CO	0	20901	1989	1993
1289008	1289008	602844128	602844128	1	UNION CARBIDE CORPORATION	0	20901	1994	1994
1289008	1289008	$<$	1289008	1	UNION CARBIDE CORPORATION	0	19400	1977	1977

Note: This table shows an example Dun & Bradstreet record. The construction of this entry follows the procedure described in Appendix B.1. Note that we are missing Dun & Bradstreet files for 1981 and 1984. We account for this by filling forward activity data in those years.

Table B.3: **Merge quality between PatentsView and Dun & Bradstreet**

Match quality				Firm resolution		
Statistic	Matched	PatentsView	Match rate	Organization	N	Cum. share
Number of assignees	150730	205655	73%	Headquarters	2375886	0.86
Number of commuting zones	679	707	96%	Single establishment	339269	0.98
Number of patents	2759380	3269141	84%	Ultimate owner	44225	1.00
Grant year						
Mean	2005	2006				
Median	2009	2009				
Application year						
Mean	2003	2003				
Median	2005	2005				

Note: This table conditions on nonwithdrawn utility patents granted between 1969 and 2021 with a single U.S.-based assignee.

of firm activity in the resulting sample. In particular, the average firm operates in 2 markets, generates 7 patents per year, and employs 11 inventors, even though the medians are substantially smaller.

Table B.4: **Summary statistics**

Statistic	Mean	Median
Market	2.09	1.00
Establishments	4.62	1.00
Patents	6.93	1.00
Inventors	10.74	2.00
D&B employment	507.69	26.00
External citations	63.30	2.00

Note: The table reports firm-level statistics from the merged PatentsView–Dun & Bradstreet data for firms with at least one patent and positive inventor employment. The data are pooled across years.

The table also highlights that the median patenting firm remains geographically narrow, with activity concentrated in a single market, whereas the mean is pulled up by a smaller set of firms with broader geographic footprints. This skewness helps motivate the emphasis in the main text on multi-market firms and the concentration of innovation within a relatively small set of geographically broad organizations.

Our main analysis uses a dataset that merges the PatentsView citations data with Dun & Bradstreet establishment data. Since Dun & Bradstreet establishment data is used to construct the National Establishment Time Series (NETS) data, we follow Hughes et al. (2021) and perform an "exact" merge with commuting zone level blocking. To give a sense of our data structure, Tables B.1 and B.2 show example records that we are trying to match. We merge records using the following procedure.

1. Using the raw Dun & Bradstreet data, we construct a dataset with the first and last year a given Dun & Bradstreet observation appears. Dun & Bradstreet observations are indexed by their `dunsno-{companyname/secondaryname}-CZ1990`.

2. Using Stata packages developed by Wasi and Flaaen (2015), we standardize the firm names in the Dun & Bradstreet data and in the patent data. For the patent data, we use the file `g_assignee_not_disambiguated.tsv.zip` from PatentsView, downloaded on December 15, 2023. The assignee file has geographic information on the assignee FIPS code, which we match to 1990 commuting zones using a crosswalk developed by the Economic Research Service (ERS).

3. We perform an exact merge on name and commuting zone between the two datasets

   (a) Due to systematic asymmetric naming in the two data sets, we hand-merge all IBM records to `dunsno 001368083`.

   (b) Using `companyname`, we match on standardized name and commuting zone using the configuration of Wasi and Flaaen (2015) suggested by Hughes et al. (2021).

   (c) Next, we remove whatever matched in step 3b and match the residual on the compressed name (where we remove all white spaces from the standardized names) and commuting zone.

   (d) Lastly, we remove whatever matched in step 3c and repeat steps 3b-3c for the `secondaryname`'s in Dun & Bradstreet.

4. Step 3 can produce many potential matches for a given patent record. Moreover, the match returns establishment-level matches, whereas we wish to aggregate patenting activity to the firm level. We do this as follows.

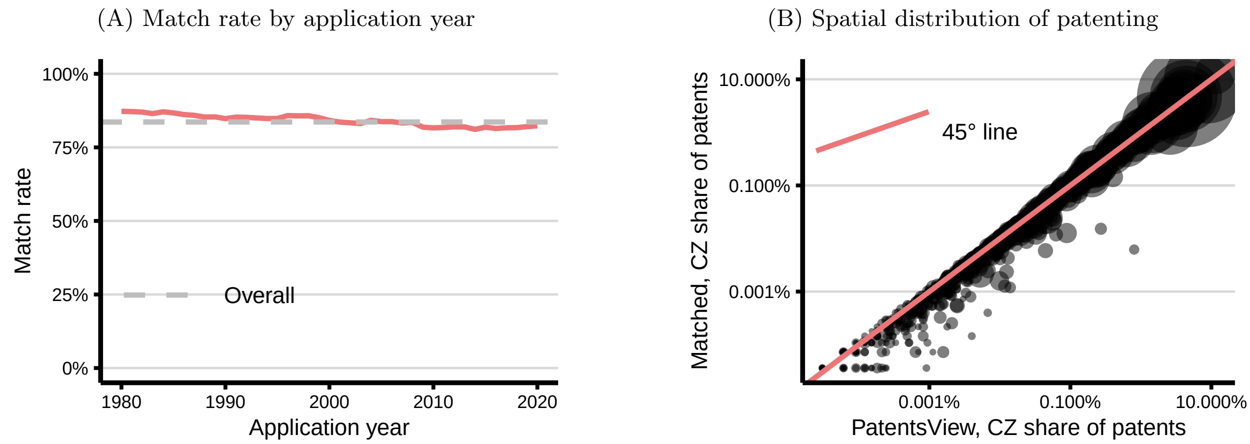
   (a) We assign codes to each match. In descending order of quality:

      i. "A" – If a patent's grant date falls within a `companyname-dunsno-CZ1990`'s first and last years of recorded activity in Dun & Bradstreet.

      ii. "B" – If a patent's grant date falls within a `dunsno-CZ1990`'s first and last years of recorded activity in Dun & Bradstreet.

      iii. "C" – If a patent's grant date falls within a `dunsno`'s first and last years of recorded activity in Dun & Bradstreet.

Figure B.1: **Match quality across time and space**



Note: The left panel shows the match rate as a function of patent application year. The grey dashed line is the overall match rate. The right panel shows the spatial distribution of patenting in the matched sample and the full PatentsView population. Each dot represents a commuting zone and its share of patenting in the matched sample and in the full PatentsView population. The sizes of the dots correspond to employment in each commuting zone. The right panel has $\log_{10}$ scaling along the x- and y-axes.

iv. "D" – If a patent does not satisfy any of the above.

v. Note that for the example records in Tables B.1 and B.2, after name standardization, there would be a single type "C" match.

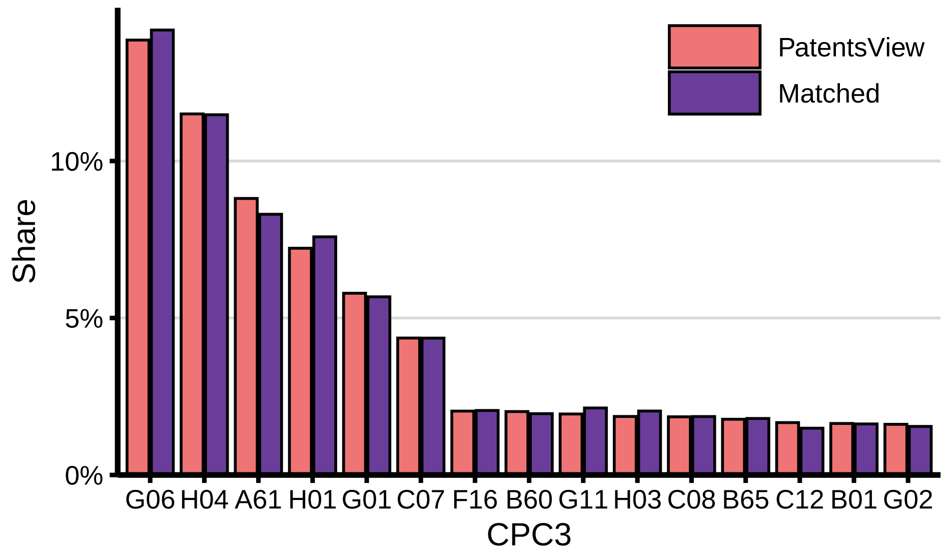
(b) We associate each match with its "nearest" record in Dun & Bradstreet. For type "A" matches, this nearest record is an exact match. For all other types, we find the record nearest to the given patent's grant date.

(c) Then we assign the patent to a unique Dun & Bradstreet identifier, which we denote by `firm_id` in what follows.

- If the patent is associated with no `hqdunsno`'s (i.e. it is only associated with `dunsno`'s), then we assign equal ownership to all unique `dunsno`'s
- If the patent is associated with a single unique `hqdunsno`, then we assign the patent to the unique `hqdunsno`
- If there are multiple unique `hqdunsno`, we refine in the following order:
  i. Multiple `dunsno` report HQ activity. Take all unique `ultdunsno`'s.
  ii. All `dunsno` report single location or branch activity. Take all unique `hqdunsno`'s.
  iii. There is a single `ultdunsno`. Take the unique `ultdunsno`.
  iv. Otherwise, take all unique `hqdunsno`.

Table B.3 and Figures B.1 and B.2 show that the merged sample remains close to the PatentsView universe along the dimensions most relevant for our analysis. Table B.3 shows that we match 84% of patents, 73% of assignee organizations, and 96% of commuting zones, while the mean and median grant and application years in the matched sample are nearly identical to those in PatentsView.

Figure B.2: **Frequency distribution of three digit CPC in PatentsView and in the matched sample**



Note: This figure compares the distribution of patent counts across the 15 three-digit CPC classes with the largest patent counts in our data in the full PatentsView sample and in the matched PatentsView–Dun & Bradstreet sample.

Figure B.1 shows that the match rate is broadly stable over time and that the spatial distribution of patenting in the matched sample closely mirrors the population distribution across commuting zones. Figure B.2 complements this evidence by showing that the technology composition of the matched sample closely tracks the PatentsView distribution across three-digit CPC classes, indicating little technological imbalance induced by the merge. Taken together, these patterns suggest that the merge preserves the temporal, technological, and geographic composition of innovative activity while delivering high overall coverage.

B.2 Allocating inventors to firms

For certain of our analyses we require a panel of inventor-firm links. The patent data do not provide this directly, but we are able to construct one as follows.

1. `g_inventor_disambiguated.tsv.zip` provides the disambiguated inventors for each of our patents as well as their self-reported locations.

2. We merge the patent ownership dataset (Appendix B.1) onto the Dun & Bradstreet data by `dunsno-year` (where `year` for the patent ownership data is the application year). Establishment activity is considered within a +2 year window around the report date (i.e. if a patent is applied for in 2008 and Dun & Bradstreet reports establishment activity any time between 2008-2010, this would be included in our merge). This is a many-to-many merge.

3. By `patent_id` and `cz1990` we merge the inventor file onto the dataset created in step 2. We

take all unique inventor_id-year-firm_id-cz1990 matches.

- If an inventor reports multiple commuting zones in a year, it may be assigned to multiple firm branches in a given year.
- We attempt to refine this match by looking at the inventor's self-reported FIPS (i.e. state-county pair) and checking for establishment activity in those counties. If this procedure enables us to eliminate a multiplicitous commuting zone, we do so; if multiple counties have establishment activity, we keep the multiple matches.

4. We residualize any remaining inventor_id-year's who could not be exact matched on commuting zone and merge this residual, by patent_id, on to the ownership file. Within an inventor_id-year, we take the inventor-firm pair with the smallest geographical distance between the inventor_id-FIPS and the firm_id-FIPS as defined using geographical centroids of the counties.

5. Any patent that remains is either associated with an inventor outside the U.S. or is unmatched in our sample (i.e. missing an inventor_id-cz1990 or a dunsno-cz1990). Thus, it cannot be matched.

B.3 Sample selection

Our analysis covers the period from 1980 to 2018. We select these years to maximize sample size while avoiding truncation issues at both ends of the data. While our patent data includes patents granted after 1976, many early-sample patents were applied for prior to 1976, creating incomplete coverage in the earliest years. Similarly, patents applied for after 2018 may not yet be granted in our data. We also exclude the post-2018 period to avoid complications from the Covid-19 pandemic, which disrupted the typical relationship between inventor location and workplace location which is important for the procedure described in Appendix B.2.

B.4 Comparing Dun & Bradstreet to the Longitudinal Business Database

Our analysis uses Dun & Bradstreet (D&B) to identify firm locations and activities. D&B is used to construct the National Establishment Time Series (NETS) data. Barnatchez et al. (2017) study how similar NETS is to the Longitudinal Business Database (LBD). While it has known limitations compared to the LBD, these concerns are largely mitigated for our specific application. Our research design is well-suited to D&B's strengths while avoiding its main weaknesses. First, as we show, patenting activity is heavily concentrated among larger establishments, where Barnatchez et al. (2017) show that NETS correlates highly with official sources (above 95% for establishment counts and above 80% for employment levels). Second, our analysis focuses mostly on the extensive margin – whether firm activity exists in a given location – rather than firm size or growth patterns, making the concerns Barnatchez et al. (2017) raise about imputation largely irrelevant to our study.

For our purposes, D&B provides comprehensive geographic coverage and establishment detail that makes it well-suited for linking to patent records and identifying the spatial distribution of innovation activity.

C Details of the matching procedure for difference-in-differences analysis

For the difference-in-differences analysis, treatment timing is based on the first year in which a patenting firm is observed with a D&B establishment in a commuting zone. We use first observed establishment presence rather than first local patenting so that expansion can precede measured patent output in the new market.

- For a treated unit ωn (i.e. where $g_{\omega n} < \infty$), consider all units $\omega n'$ where $g_{\omega n'} = \infty$. These $\omega n'$ are potential control units. We allow for potential control units to be repeated within a multi-establishment firm, across treated units (i.e. we will match with replacement).
- We require both treated and potential control units to have received positive citations at least twice for $\tau \equiv t - g_{\omega n} \in \{-10, \ldots, -1\}$.
- Let $y_{\omega n\tau}$ denote an outcome of interest for firm × CZ1990 ωn at relative time τ.
- **Matching on pre-treatment outcomes, *changes***

$$\begin{aligned} \texttt{SLOPE}_{\omega n} &= \frac{\sum_{\tau=-10}^{-6} y_{\omega n\tau}}{5} - \frac{\sum_{\tau=-5}^{-1} y_{\omega n\tau}}{5} \\ \texttt{CRIT1}_{\omega nn'} &= \frac{\texttt{SLOPE}_{\omega n} - \texttt{SLOPE}_{\omega n'}}{0.5\left(|\texttt{SLOPE}_{\omega n}| + |\texttt{SLOPE}_{\omega n'}|\right)} \end{aligned}$$

  - If $\texttt{SLOPE}_{\omega n}$ and $\texttt{SLOPE}_{\omega n'}$ have different **signs**, then remove these potential matches.
- **Matching on pre-treatment outcomes, *levels***

$$\texttt{CRIT2}_{\omega nn'} = \frac{1}{|K|} \sum_{\tau \in K} \left(\log y_{\omega n\tau} - \log y_{\omega n'\tau}\right)$$

  - K is a subset of $\{-10, ..., -1\}$ where both $y_{\omega n\tau}$ and $y_{\omega n'\tau}$ are strictly positive
- **Matching on pre-treatment covariates**
  1. **Employment shares** – broad employment composition in commuting zones n and n'
  2. **Business Dynamics** – Employment to firm and establishment to firm ratios for two cohorts of firm types:
     (a) Large firms
     (b) Overall

3. **Demographics**
   (a) Log income per capita
   (b) Log employment
   (c) Ratio of college-educated to noncollege-educated workforces

$\texttt{CRIT3}_{\omega n n'}$ is equal to the mean of the Davis, Haltiwanger and Schuh (1998) (DHS) differences between n and n' for all of the listed covariates over $\tau \in \{-10, \ldots, -1\}$.

- **Select control unit for each treated unit**
  - Take $n'_{\omega}(n) = \arg\min_{n'} \texttt{CRIT}_{\omega n n'}$ (break ties randomly)

$$\texttt{CRIT}_{\omega n n'} = \frac{1}{3}\left(\texttt{CRIT1}_{\omega n n'} + \texttt{CRIT2}_{\omega n n'} + \texttt{CRIT3}_{\omega n n'}\right).$$

D Robustness

D.1 Alternative specifications

Table D.1 reports the underlying coefficients for two of the event studies in the main text. The estimates make clear that the main results combine little evidence of pre-trends with a gradual and persistent rise in both external citations and patenting efficiency after entry.

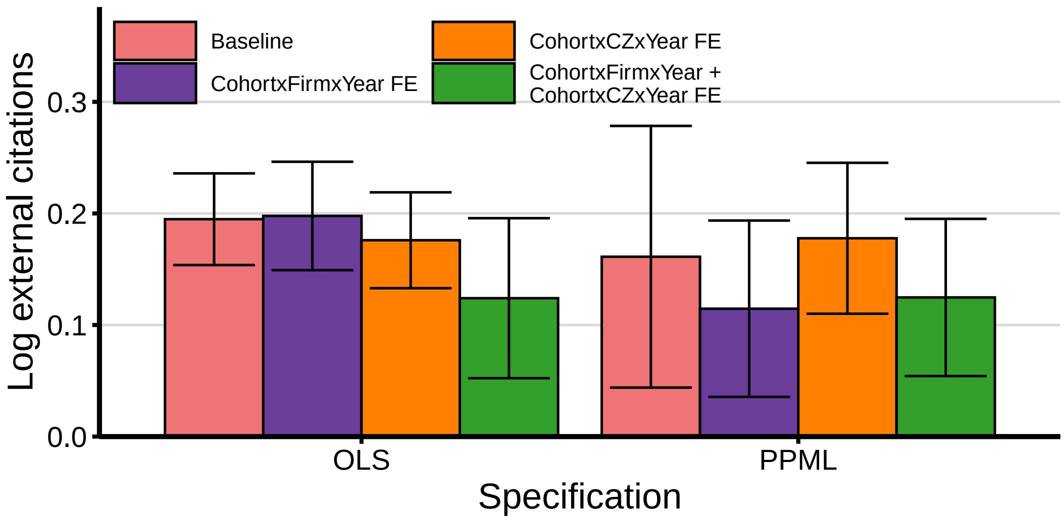
Figure D.1 compresses the dynamic evidence into a static difference-in-differences specification and shows that the external-citation effect remains positive under OLS and PPML across the full set of fixed-effect structures. The most saturated specification is somewhat smaller and less precise, but the basic conclusion that entry raises local citations is unchanged.

Figure D.2 repeats this exercise for the within-CPC outcomes. Same-CPC citations and patenting efficiency remain positive under both OLS and PPML, and the specifications with cohort-firm-year controls stay close to the baseline estimates.

Figure D.3 excludes firms linked to the focal firm through observed supply-chain relationships. The citation and patenting-efficiency responses remain positive and look similar to the benchmark event studies, suggesting that the main results are not driven by vertical relationships between the entrant and local firms.

Figure D.4 imposes the stricter citation definition discussed in the main text. Across the baseline, firm-year, and commuting-zone-year specifications, the post-entry increase in citations remains clear and persistent, with little evidence of pre-trends.

Figure D.1: **External citation estimates across static difference-in-differences specifications.**

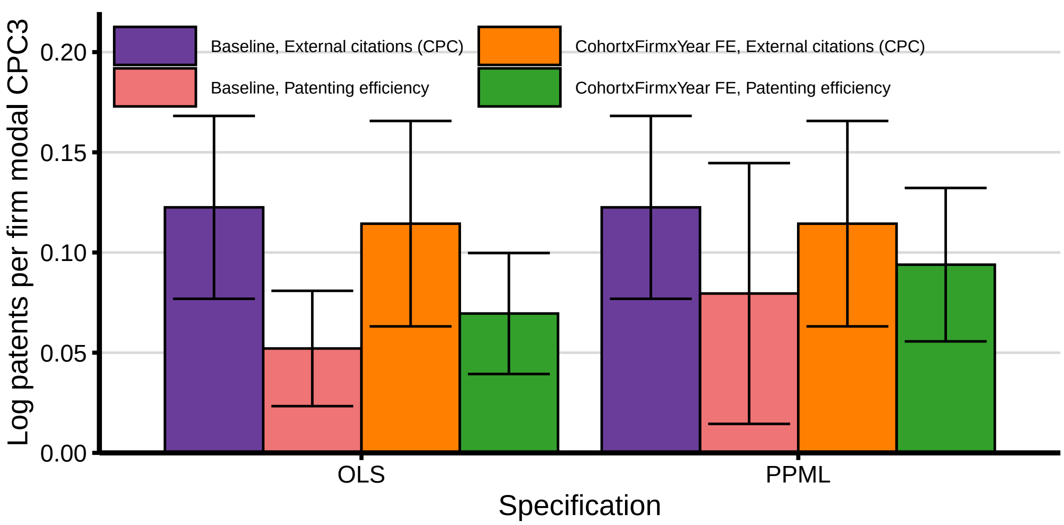


Note: Figure reports the results of the static difference-in-differences regression:

$$y_{\omega n t} = \mathbf{FE}_{\tilde{g}_{\omega n}\omega n} + \mathbf{FE}_{\tilde{g}_{\omega n}t\cdot} + \delta\mathbb{I}\{t \geq g_{\omega n}\} + \varepsilon_{\omega n t}$$

where $\cdot$ denotes the additional fixed effects listed in the figure. We report specifications with baseline fixed effects, cohort×firm×year fixed effects, cohort×commuting zone×year fixed effects, and both sets of controls jointly. OLS uses log external citations and excludes zeros, while PPML uses the citation count and includes zeros. Error bars represent 95% confidence intervals. Standard errors are clustered at the firm×commuting zone×cohort level.

Figure D.2: **Within-CPC outcomes across static difference-in-differences specifications.**



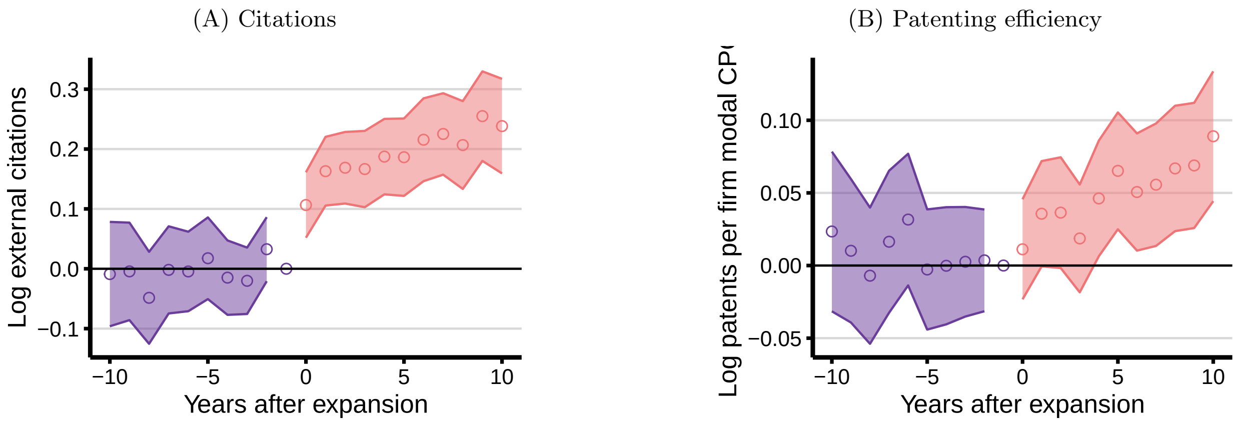
Note: Figure reports static difference-in-differences estimates of the post-entry effect for within-CPC external citations and patenting efficiency under OLS and PPML. For each estimator, we show the baseline specification and the version with cohort×firm×year fixed effects. OLS uses log outcomes, while PPML uses the raw outcomes and includes zeros. Error bars represent 95% confidence intervals. Standard errors are clustered at the firm×commuting zone×cohort level.

Table D.1: **Estimation results for main-text event studies**

τ	External citations	Patenting efficiency
-10	-0.02	0.02
	(0.04)	(0.03)
-9	-0.01	0.02
	(0.04)	(0.02)
-8	-0.06	-0.00
	(0.04)	(0.02)
-7	-0.02	0.01
	(0.04)	(0.02)
-6	-0.02	0.03
	(0.03)	(0.02)
-5	0.02	-0.01
	(0.03)	(0.02)
-4	-0.02	-0.00
	(0.03)	(0.02)
-3	-0.03	0.00
	(0.03)	(0.02)
-2	0.03	-0.00
	(0.03)	(0.02)
0	0.11	0.02
	(0.03)	(0.02)
1	0.16	0.04
	(0.03)	(0.02)
2	0.17	0.05
	(0.03)	(0.02)
3	0.16	0.03
	(0.03)	(0.02)
4	0.18	0.06
	(0.03)	(0.02)
5	0.18	0.08
	(0.03)	(0.02)
6	0.20	0.06
	(0.03)	(0.02)
7	0.22	0.07
	(0.03)	(0.02)
8	0.20	0.07
	(0.04)	(0.02)
9	0.24	0.08
	(0.04)	(0.02)
10	0.23	0.10
	(0.04)	(0.02)
Adjusted R^2	0.47	0.70
Wald test pre-trend p-value	0.50	0.73
N	180797	255493
Number of treatment units	5789	5747

Note: This table reports the regression loadings from Panels 3A and 4C. Standard errors are clustered at the firm $\times$ commuting zone $\times$ cohort ($\omega \times n \times \tilde{g}_{\omega n}$) level.

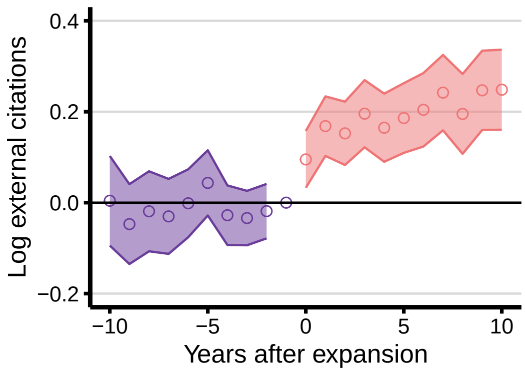
Figure D.3: **Event studies excluding firms that are suppliers or customers of one another.**



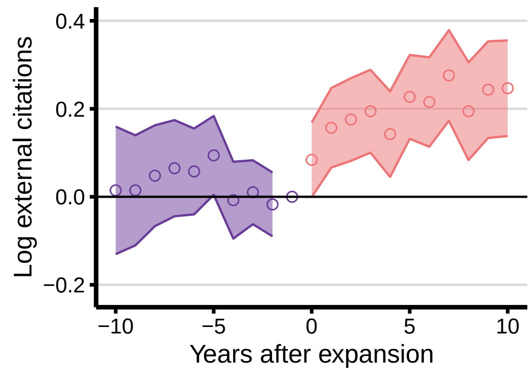
Note: The figure repeats the event study regression (13) for external citations and patenting efficiency after removing citations, patents, and firms linked to the focal firm through an observed supply-chain relationship. We identify these links by crosswalking our D&B firm identifiers to Compustat GVKEYs using our D&B-patent bridge and the Compustat-patent bridge of Bloom et al. (2013), and then using Compustat Segment customer-supplier links (Cen et al., 2017; Cohen and Frazzini, 2008). Error bands represent 95% confidence intervals. Wald tests of the joint equality of the pre-treatment coefficients to zero are not rejected ($p = 0.65$ and 0.73, respectively).

Figure D.4: **External citation event studies after expansion into a new market, using a more stringent definition of external citations.**

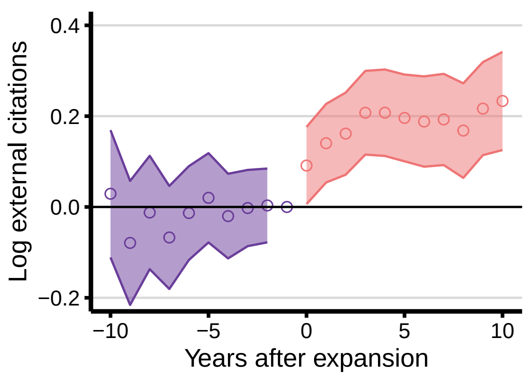
(A) Baseline specification



(B) With firm×year FE



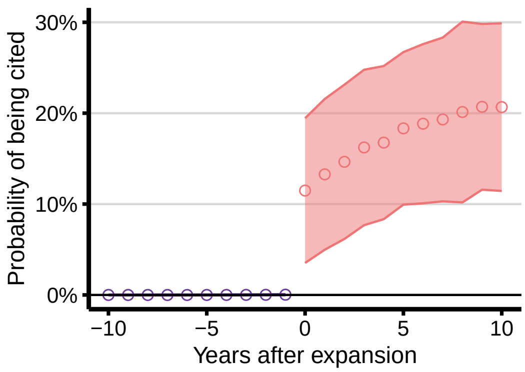
(C) With commuting zone×year FE



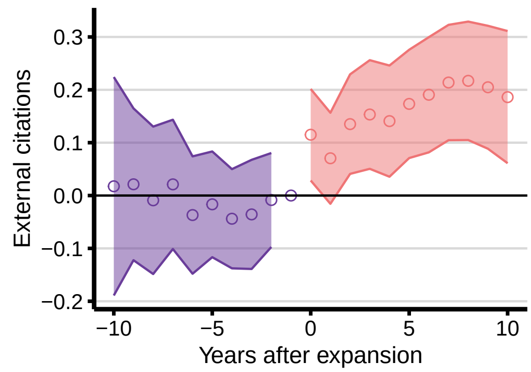
Note: The figure shows the event study regression (13), where the dependent variable is log external citations, with two additional restrictions. (1) We count only citations to patents of firm ω invented by ω's inventors located outside the expansion commuting zone. (2) We count only citations made by patents of other firms $\omega' \neq \omega$ whose inventors never patent for firm ω. Error bands represent 95% confidence intervals. Standard errors are clustered at the firm–commuting zone–cohort level. Panel D.4A shows the event study under the baseline specification (Equation 13); Panel D.4B adds a firm-year-cohort control; and Panel D.4C adds a commuting zone-year-cohort control. Wald tests of the joint equality of the pre-treatment coefficients to zero are not rejected ($p = 0.49$, 0.51, and 0.89, respectively).

Figure D.5: **Event studies, alternative estimators.**

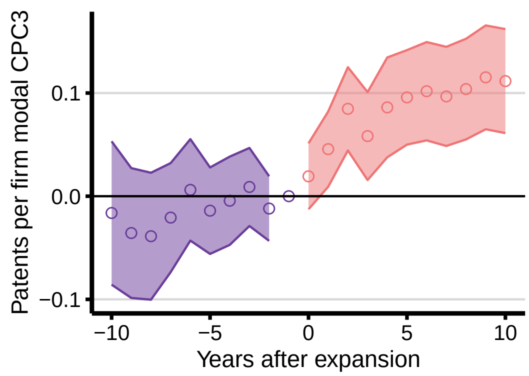
(A) Probability of receiving a citation after expansion



(B) PPML, citations



(C) PPML, patenting efficiency



Note: Panel D.5A presents event study coefficients from the specification in Equation (13), capturing the response of the probability that a firm receives at least one external citation in a commuting zone (CZ) following its entry into that CZ. Here, we study an alternative dependent variable equal to an indicator function for positive citations (i.e. a linear probability model, LPM) with a different matching structure based on synthetic difference-in-differences (Arkhangelsky et al., 2021; Ben-Michael et al., 2022). Figure D.6 presents density plots of the bootstrap estimates for each treatment-time dummy. Panels D.5B and D.5C show our event studies estimated using pseudo-Poisson Maximum Likelihood (PPML) (Bergé, 2018; Correia et al., 2020). These panels include a cohort-firm-year fixed effect. For Panel D.5A, error bands are the empirical 2.5^{th} and 97.5^{th} percentiles of the bootstrap distributions. For Panels D.5B and D.5C, error bands represent 95% confidence intervals. The Wald test of joint equality of the pre-treatment coefficients to zero is not rejected for Panels D.5B and D.5C ($p = 0.97$ and 0.60, respectively).

D.2 Alternative estimators

Panel D.5A applies the synthetic difference-in-differences estimator of Ben-Michael et al. (2022) to an extensive-margin outcome equal to an indicator for receiving any external citation. We make this switch because the package does not permit matching with missing outcome observations, whereas our baseline log specification drops zero-citation cells. The full extensive-margin panel is also too large for the package, so we estimate this specification on bootstrap draws from the main regression sample and report the corresponding coefficient distributions in Figure D.6.

Table D.2: **Estimation on Million Dollar Plant expansions**

	External		Patenting efficiency	
	OLS	PPML	OLS	PPML
$\mathbb{I}\{t \geq g_{\omega n}\}$	0.44	0.52	0.13	0.18
	(0.19)	(0.16)	(0.12)	(0.12)
Adjusted R^2	0.59	0.74	0.72	0.35
N	1065	1792	834	1416
Number of treatment units	40	40	29	29

Note: This table shows the results of the regression:

$$\log y_{\omega nt} = \alpha_n + \delta t + c_1 \mathbb{I}\{t \geq \tilde{g}_{\omega n}\} + c_2 \mathbb{I}\{g_{\omega n} < \infty\} + c_3 \mathbb{I}\{t \geq g_{\omega n}\} + \varepsilon_{\omega nt}$$

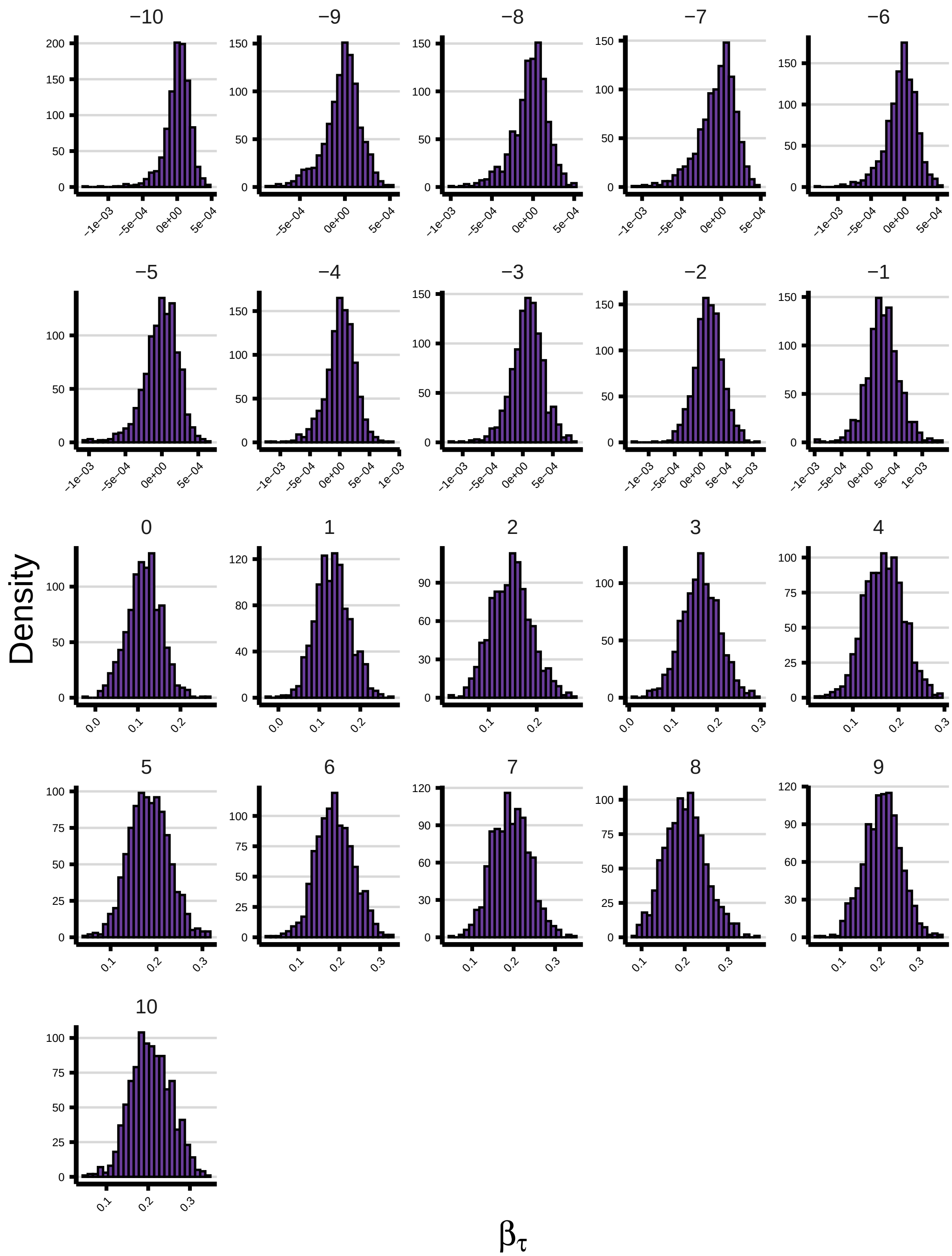
where n indexes a county. Columns 1–2 report external citations and Columns 3–4 report patenting efficiency on the million-dollar plant sample of Greenstone et al. (2010). Within each outcome, we report OLS and PPML estimates of the post-entry effect.

Panels D.5A, D.5B, and D.5C compare the baseline event study to alternative estimators. Panel D.5A shows that the probability of receiving any citation rises after entry in the synthetic difference-in-differences extensive-margin specification, with little evidence of pre-treatment movement. The PPML event studies for citations and patenting efficiency likewise closely track the benchmark results, confirming that the findings are not driven by the log transformation or the treatment of zeros.

Table D.2 reports estimates on the much smaller sample of million-dollar plant expansions. The coefficients are positive in all four columns and sizeable for external citations, but the sample is small enough that these estimates are naturally less precise than the main results.

Figure D.6 plots the bootstrap distributions underlying Panel D.5A. The pre-treatment coefficients remain centered near zero, and the post-treatment coefficients shift steadily to the right of zero. The distributions are also close to symmetric.

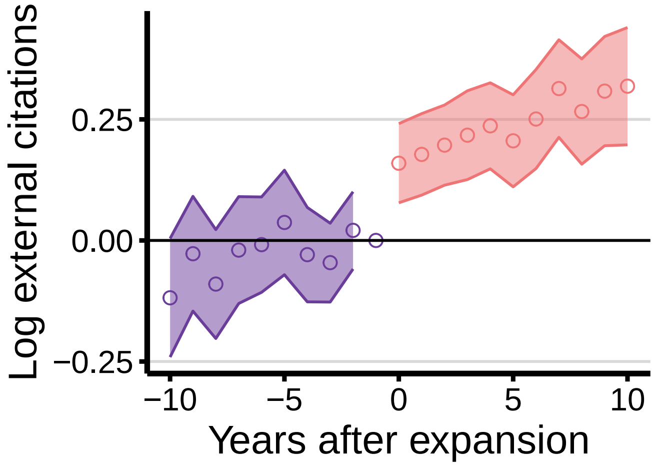
Figure D.6: **Bootstrap distribution of synthetic difference-in-differences estimates**



Note: This figure shows the empirical density of the bootstrap estimates for each treatment-time coefficient in the synthetic difference-in-differences specification (Arkhangelsky et al., 2021; Ben-Michael et al., 2022). Each panel corresponds to one relative-time coefficient. Estimates are computed from 1,000 bootstrap replicates on random 0.5% samples of the main regression data, sampling over $\omega \times n$ units without replacement.

D.3 Persistent treatment status

Figure D.7: **External citation event study excluding temporary treated units**



Note: The figure repeats Equation (13) for log external citations after excluding treated units that exit the destination commuting zone within ten years of entry. Error bands represent 95% confidence intervals. The Wald test of joint equality of the pre-treatment coefficients to zero is not rejected ($p = 0.26$). Standard errors are clustered at the firm × commuting zone × cohort ($\omega \times n \times \tilde{g}_{\omega n}$) level.

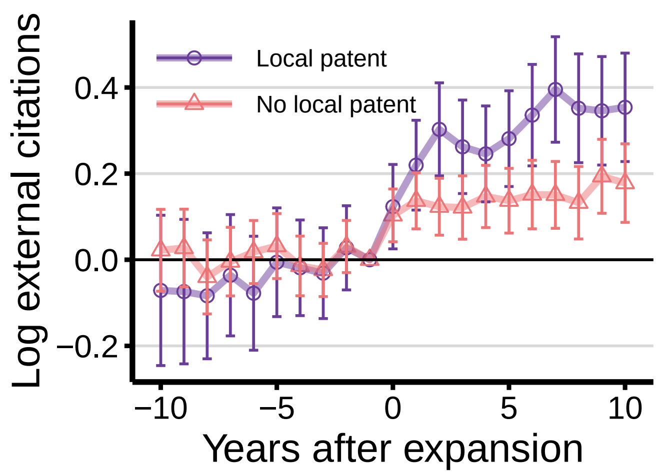
Figure D.7 shows that the citation effect remains positive and is highly similar after excluding treated markets that are later exited.

D.4 Additional heterogeneity analysis

Figure D.8 splits expansions by whether the entrant patents locally within five years of arrival. Spillovers are positive in both groups, but they are clearly larger when the new location quickly becomes an active patenting site, which is consistent with local innovative activity amplifying the knowledge flows captured in the baseline estimates.

Figure D.9 instead partitions the sample by whether the firm's technology matches the destination commuting zone's dominant two-digit CPC. The static and dynamic estimates are similar across the two groups, with if anything slightly larger point estimates for non-matching expansions. This is consistent with Figure 2B in the main text, which shows that firms tend to replicate similar innovation activities across markets rather than tailor them tightly to the destination market's existing technology mix.

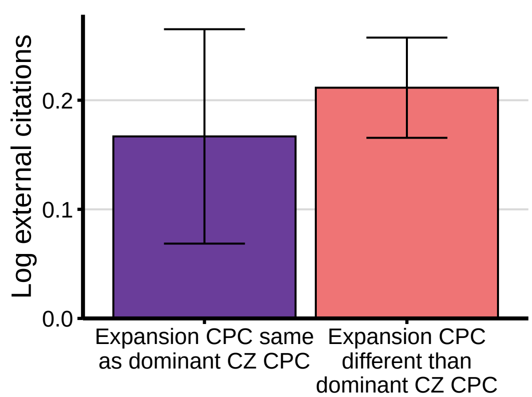
Figure D.8: **Heterogeneity by local patenting activity**



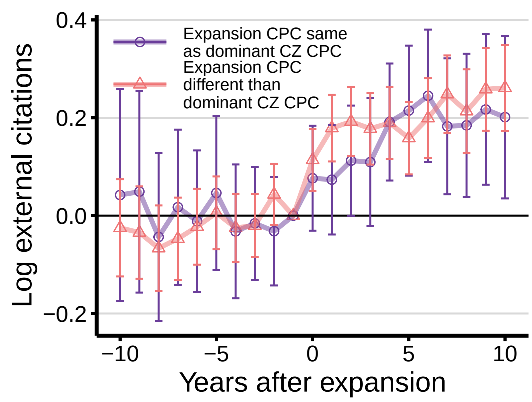
Note: The figure reports event study coefficients for external citations separately by whether the expanding firm patents locally within five years of entry. Error bars represent 95% confidence intervals. Wald tests of joint equality of the pre-treatment coefficients to zero are not rejected ($p = 0.89$ and 0.59, respectively). Standard errors are clustered at the firm × commuting zone × cohort ($\omega \times n \times \tilde{g}_{\omega n}$) level.

Figure D.9: **Heterogeneity by technology-class match between the expanding firm and the destination commuting zone**

(A) Static difference-in-differences



(B) Dynamic difference-in-differences



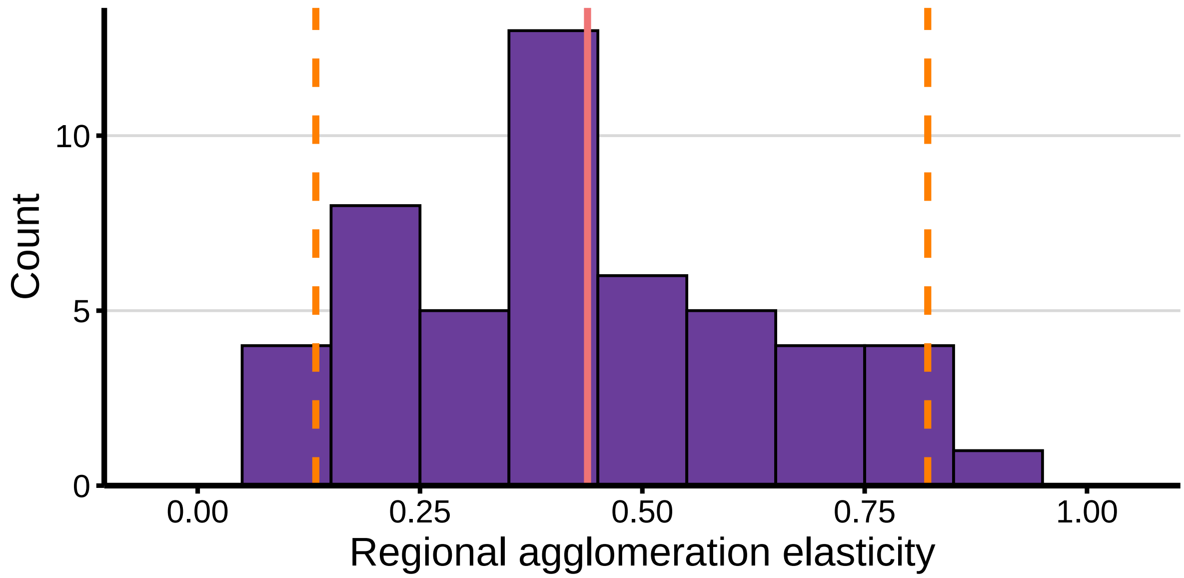
Note: We use two-digit CPCs in this analysis. Panel D.9A reports static difference-in-differences estimates of the post-entry effect, and Panel D.9B reports the corresponding dynamic event-study coefficients. Error bars represent 95% confidence intervals based on standard errors clustered at the firm × commuting zone × cohort ($\omega \times n \times \tilde{g}_{\omega n}$) level. For Panel D.9B, the Wald test of joint equality of the pre-treatment coefficients to zero is not rejected ($p = 0.96$ and 0.44 for the blue and red series, respectively).

E Additional details on the estimation

Figure E.1 plots the empirical distribution of split-sample estimates of the regional agglomeration elasticity from the bootstrap exercise.

Figure E.2 plots the bootstrap distributions of the event-study estimates of β under the baseline,

Figure E.1: **Empirical distribution of regional agglomeration elasticity estimates from bootstrap.**



Note: The figure shows the empirical distribution of split-sample estimates from regional-level regression (16). For commuting zone-year cells with more than four active firms, we randomly partition firms into two groups: one used to compute patents per firm and one used to compute market-level inventor employment. We repeat this procedure 50 times and re-estimate the regression in each iteration. The dashed orange lines represent empirical 2.5^{th} and 97.5^{th} percentiles; the solid red line represents the mean.

cohort×firm×year, and cohort×CZ×year specifications.

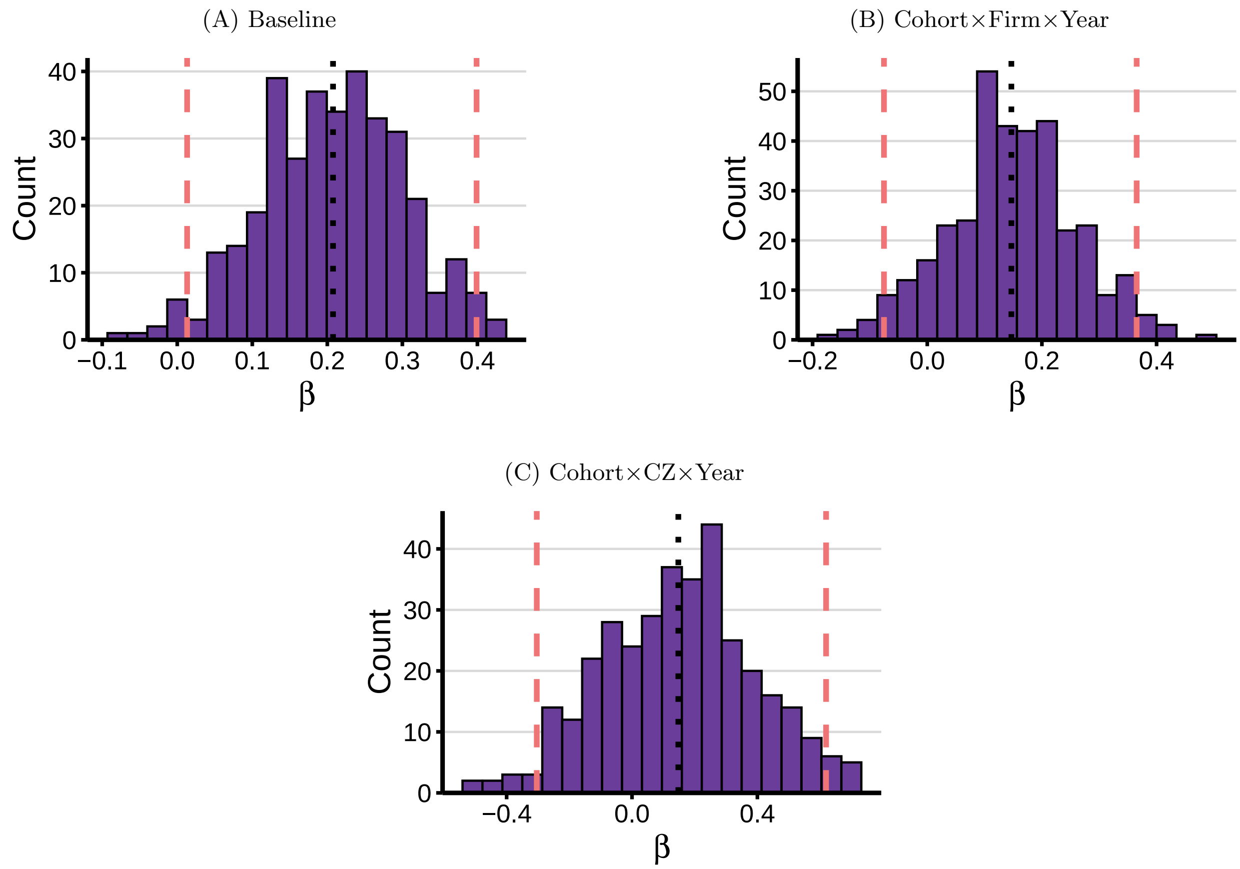
Figure E.3 plots binscatters of patenting and citations against labor input when market employment is measured using lagged inventors, with one panel for additive fixed effects and one for the full fixed-effect interaction.

Figure E.4 plots analogous binscatters using Dun & Bradstreet employment, with panels for additive fixed effects, full fixed-effect interactions, and versions that exclude markets with at most 10 employees.

E.1 Additional details on the estimation of α, β, η, and θ

For estimation of all structural parameters, we consider the period 1980-2018. For estimation of the key structural parameters, we use annual commuting-zone-year and firm-commuting-zone-year observations over 1980–2018. The regional moments are estimated at the CZ-year level, while the firm-level moments exploit firm-CZ-year variation.

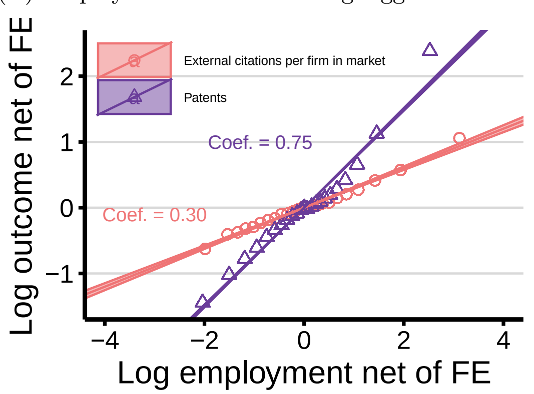
Figure E.2: **Empirical distributions of β from event study bootstrap.**



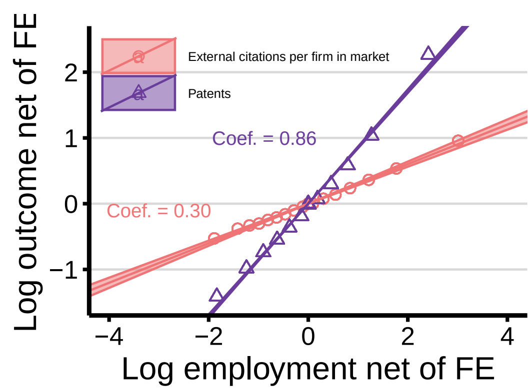
Note: This figure plots the bootstrap distributions of the event-study estimates of β reported in columns (4)–(6) of Table 5. Panel (a) is the baseline specification, panel (b) adds cohort×firm×year fixed effects, and panel (c) adds cohort×CZ×year fixed effects. The estimates are constructed from the approximation in Equation (24). The black short-dashed vertical line marks the mean, and the red longer-dashed vertical lines mark the empirical 2.5^{th} and 97.5^{th} percentiles.

Figure E.3: **Co-movement of patenting and citations with labor input, lagged inventors.**

(A) Employment in market using lagged inventors



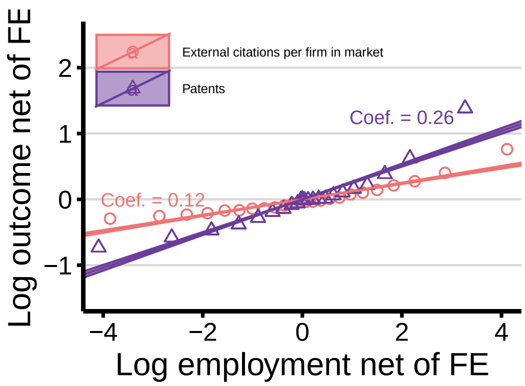
(B) Employment in market using lagged inventors, full fixed effect interaction



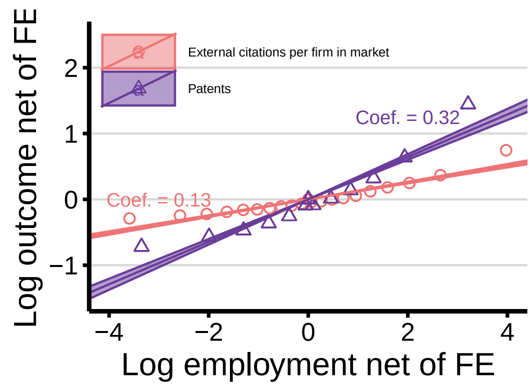
Note: The plot shows the results of regression (19) and (20) using lagged inventor employment. Panel E.3A shows the regression net of additive firm, time, and commuting zone fixed effects. Panel E.3B shows the regression net of firm×year and market×year fixed effects. Dots are binscattered into 25 equal-sized bins, net of fixed effects. Error bands are 99% confidence bands. Standard errors are clustered at the firm×market-level.

Figure E.4: **Co-movement of patenting and citations with labor input, Dun & Bradstreet employment.**

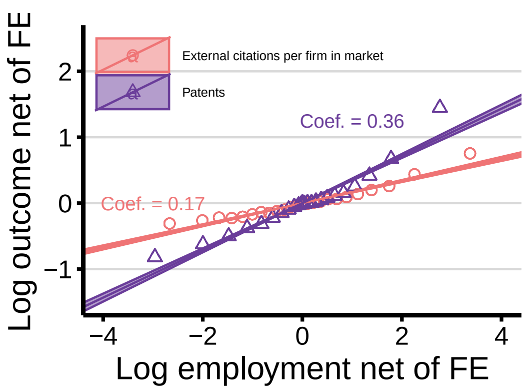
(A) Employment in market using D&B employment



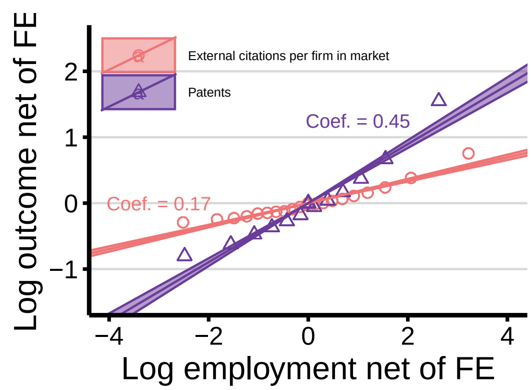
(B) Employment in market using D&B employment, full fixed effect interaction



(C) Employment in market using D&B employment, excluding markets with ≤ 10 employees



(D) Employment in market using D&B employment, full fixed effect interaction, excluding markets with ≤ 10 employees



Note: The plot shows the results of regression (19) and (20). Panel E.4A shows the regression net of additive firm, time, and commuting zone fixed effects using Dun & Bradstreet employment. Panel E.4B shows the regression net of firm×year and market×year fixed effects using Dun & Bradstreet employment. Panel E.4C shows the regression net of additive firm, time, and commuting zone fixed effects using Dun & Bradstreet employment for firms with greater than 10 employees in a given market. Panel E.4D shows the regression net of firm×year and market×year fixed effects using Dun & Bradstreet employment for firms with greater than 10 employees in a given market. Dots are binscattered into 25 equal-sized bins, net of fixed effects. Error bands are 99% confidence bands. Standard errors are clustered at the firm×market-level.

E.2 Additional details on the model inversion and estimation

E.2.1 Location fundamentals

For $\{Z_n, \bar{K}_n, B_{ni}, B_{np}\}_n$, we obtain closed-form expressions from first-order conditions of the firms and households that identify these parameters up to scale.

Production employment and wages identify $\{Z_n\}$. The monopolist producer of product line j solves the following cost-minimization problem:

$$\min_{L_{jnt}} -\sum_n w_{np} L_{jnt} \quad \text{s.t.} \quad 1 = A_{jt} \left(\sum_n (Z_n L_{jnt})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

FOC's imply that:

$$L_{jnt} = \left(\frac{w_{np}}{w_{n'p}} \right)^{-\sigma} \left(\frac{Z_{n'}}{Z_n} \right)^{1-\sigma} L_{jn't}$$

Plugging back into the constraint we see that:

$$Z_{n'} \propto w_{n'p}^{\frac{\sigma}{\sigma-1}} L_{jn't}^{\frac{1}{\sigma-1}}, \quad \forall n', j$$

In a like fashion, innovation employment and wages identify $\{\bar{K}_n\}$. First, aggregating firm-level innovation hiring in market n yields:

$$\begin{aligned}
L_{nit} &= M \int \chi_n(z) \left(\ell_{nt}(z) + \bar{f}_{nt}(z) \right) d\Psi(z) \\
&= M \int f_{max}^{-\theta} \left(\frac{1-\eta}{\eta} \right)^{\theta} \left(\ell_{nt}(z)^{\theta+1} + \frac{\theta}{\theta+1} \frac{1-\eta}{\eta} \ell_{nt}(z)^{\theta+1} \right) d\Psi(z) \\
&= M f_{max}^{-\theta} \left(\frac{1-\eta}{\eta} \right)^{\theta} \left(1 + \frac{\theta}{\theta+1} \frac{1-\eta}{\eta} \right) \left(\frac{\eta K_{nt} \lambda_0 V_t}{w_{nit}} \right)^{\frac{\theta+1}{1-\eta}} \int z^{\theta+1} d\Psi(z)
\end{aligned}$$

From the above expression it is clear that $L_{nit} \propto (K_{nt}/w_{nit})^{(\theta+1)/(1-\eta)}$

Note that:

$$\begin{aligned}
K_{nt} &= \bar{K}_n \left(M \int \chi_n(z) z^{1-\beta} \ell_{nt}(z)^{\beta} d\Psi(z) \right)^{\alpha} \\
&= \bar{K}_n \left(M f_{max}^{-\theta} \left(\frac{1-\eta}{\eta} \right)^{\theta} \left(\frac{\eta K_{nt} \lambda_0 V_t}{w_{nit}} \right)^{\frac{\theta+\beta}{1-\eta}} \int z^{\theta+\beta} d\Psi(z) \right)^{\alpha}
\end{aligned}$$

Thus:

$$K_{nt} \propto \bar{K}_n L_{nit}^{\frac{\alpha(\theta+\beta)}{1+\theta}}$$

Combining with above we get:

$$\bar{K}_n \propto w_{nit} L_{nit}^{\frac{1-\eta-\alpha(\theta+\beta)}{\theta+1}}$$

Amenities are inverted using the gravity structure implied by Gumbel preference shocks. Cobb-

Douglas preferences over housing and consumption give rise to the following expression for rents in market n:

$$R_n = \varpi\varsigma \sum_{s\in\{p,i\}} w_{ns} L_{ns}$$

Note that:

$$L_{ns} \propto B_{ns} \left(R_n^{-\varpi} w_{ns}\right)^{\epsilon}$$

due to Gumbel distribution for $\{\varepsilon_{ns}\}_{ns}$ (see main text). This easily admits an expression for $\{B_{ns}\}_{s\in\{p,i\}}$:

$$B_{ns} \propto L_{ns} \left(\frac{w_{ns}}{\left(\sum_{s'\in\{p,i\}} w_{ns'} L_{ns'}\right)^{\varpi}}\right)^{-\epsilon}$$

E.2.2 Model parameters

The remaining parameters $\{\gamma, \lambda_0, f_{max}, \bar{z}, f_e\}$ are common across locations and are calibrated within the solution to the balanced growth path equilibrium. We do this by first posing a guess for the equilibrium − which is characterized by de-trended wages in both sectors in all locations; de-trended rents in all locations; innovation efficiencies (given $\bar{K}_n$ from above) in all locations; the mass of active firms; and the de-trended market value of a firm under a set of initial parameters: $\{\gamma, \lambda_0, f_{max}, \bar{z}, f_e\}_0$. We embed in this equilibrium a set of moments from the data that we use to update the equilibrium, which then implies new values for parameters (described in further detail below). We iterate until the equilibrium vector converges to an arbitrary precision.

For γ note that national income accounting implies that:

$$Y = \rho V + M f_e + \sum_{s,n} w_{ns} L_{ns}$$

The quality ladder structure of the model plus Cobb-Douglas aggregation of intermediate varieties further implies that net profits of the firm equal $\frac{\gamma-1}{\gamma}Y$. Since revenues equal Y, this leaves $\frac{1}{\gamma}Y$ units left to compensate labor. Therefore $\sum_n w_{np}L_{np} = \frac{1}{\gamma}Y$. Combining these two expressions yields:

$$\begin{aligned} Y &= M f_e + \rho V + \frac{1}{\gamma} Y \left(1 + \frac{\sum_n w_{ni} L_{ni}}{\sum_n w_{np} L_{np}}\right) \\ \gamma &= \frac{1 + \frac{\sum_n w_{ni} L_{ni}}{\sum_n w_{np} L_{np}}}{1 - \frac{M f_e + \rho V}{Y}} \end{aligned} \tag{53}$$

where $\frac{\sum_n w_{ni} L_{ni}}{\sum_n w_{np} L_{np}}$ is the ratio of labor income in the innovation sector to labor income in the production sector and is taken from data.

We discipline λ_0 using the growth rate. Note that $g = \bar{\lambda} \log \gamma$ (Proposition 4). The expression for $\bar{\lambda}$ is messy, but λ_0 is easily separable. From Proposition 4:

$$\bar{\lambda} = M\lambda_0 \int \sum_n K_n \chi_n(z) z^{1-\eta} \ell_n(z)^\eta d\Psi(z)$$

Note that:

$$\chi_n(z) z^{1-\eta} \ell_n(z)^\eta = \underbrace{f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^\theta}_{\mathcal{C}_1} \left(\frac{\eta K_{nt} \lambda_0 V_t}{w_{nit}}\right)^{\frac{\theta+\eta}{1-\eta}} \times \overbrace{\int z^{1+\theta} d\Psi(z)}^{\bar{\mathcal{Z}}}$$

So that:

$$\lambda_0 = \left[\frac{M}{\bar{\lambda}} \sum_n K_n^{\frac{\theta+1}{1-\eta}} \mathcal{C}_1 \left(\frac{\eta V_t}{w_{nit}}\right)^{\frac{\theta+\eta}{1-\eta}} \bar{\mathcal{Z}}\right]^{\frac{\eta-1}{1+\theta}} \tag{54}$$

We get g from data.

f_{max} is a scale parameter for the draws of fixed costs; it is thus intimately tied to firms' spatial scope. The measure of active firms is given by:

$$M\left(1 - \int \prod_n (1 - \chi_n(z)) d\Psi(z)\right) = M\bar{\chi}.$$

The number of establishments (markets) that are active is given by:

$$M \int \sum_n \chi_n(z) d\Psi(z)$$

Therefore the average number of markets per firm is given by:

$$\frac{\#\text{ Markets}}{\#\text{ Firms}} = \frac{\int \sum_n \chi_n(z) d\Psi(z)}{\bar{\chi}} = \frac{f_{max}^{-\theta} \left(\frac{1-\eta}{\eta}\right)^\theta (\eta\lambda_0 V_t)^{\frac{\theta}{1-\eta}} \sum_n \left(\frac{K_{nt}}{w_{ni}}\right)^{\frac{\theta}{1-\eta}} \int z^\theta d\Psi(z)}{\bar{\chi}}$$

yielding:

$$f_{max} = \left[\frac{\left(\frac{1-\eta}{\eta}\right)^\theta (\eta\lambda_0 V_t)^{\frac{\theta}{1-\eta}} \sum_n \left(\frac{K_{nt}}{w_{ni}}\right)^{\frac{\theta}{1-\eta}} \int z^\theta d\Psi(z)}{\bar{\chi} \frac{\#\text{ Markets}}{\#\text{ Firms}}}\right]^{\frac{1}{\theta}} \tag{55}$$

where $\frac{\#\text{ Markets}}{\#\text{ Firms}}$ comes from data.

We adopt a simple distribution for $\Psi(z)$. With probability 0.9, a firm draws $z = \underline{z} = 1$; with complementary probability 0.1 it draws $z = \bar{z}$. We set $\bar{z}$ such that 70% of innovations are accounted for by the $z = \bar{z}$ (i.e. the top 10%) firms, which matches our data.[30]

[30] Our patent data roughly obeys the Pareto principle, consistent with evidence reported elsewhere that innovation is a fat-tailed phenomenon (Scherer and Harhoff, 2000).

Lastly, one can verify that:

$$f_e = \frac{\bar{\lambda} V_t (1-\eta)}{(\theta+1) M} \tag{56}$$

which completes specification of our model inversion equations. The key inversion updates used below are summarized by Equations (53), (54), (55), and (56).

To complete the description of the model inversion, we provide a rough outline of how we jointly solve and calibrate model parameters.

1. Invert $\{Z_n, \bar{K}_n, B_{ni}, B_{np}\}_n$ according to section E.2.1.
2. Guess $\{\gamma, \lambda_0, f_{max}, \bar{z}, f_e\}$ and $\{\{w_{ni}, w_{np}, R_n, K_n\}_n, M, V\}$
3. Using the guess, solve for $\{L_{np}, L_{ni}\}_n$ and update $\{R_n\}_n$ according to:

   $$R_n^{(k+1)} = \Delta R_n^k + (1-\Delta) R_n^k$$

   where Δ is some number close to but less than 1
4. Given the guess for γ and $\{L_{np}\}_n$ update $\{w_{np}\}_n$.
5. Given the guess for $\{\{K_n\}_n, \lambda_0, V, M\}$ and $\{L_{ni}\}_n$ update $\{w_{ni}\}_n$.
6. Given the guess for $\{M, \lambda_0, V\}$ and the new values for $\{w_{ni}\}$ update $\{K_n\}_n$.
7. Given the guess for f_{max}, $\{L_{ni}\}_n$, and data on the average firm size, update M.
8. Update f_e according to (56).
9. Using guess for $\gamma, \bar{\lambda}$, and $\{L_{np}\}_n$ update V.
10. Using guess for γ and data on the growth rate, update $\bar{\lambda}$.
11. Update λ_0 using (54).
12. Update γ using (53).
13. Update f_{max} using (55) and dampener Δ.
14. Check for convergence of $\{\{w_{ni}, w_{np}, R_n, K_n\}_n, M, V\}$; if converged then stop; otherwise repeat.