Innovation Spillovers across U.S. Tech Clusters

Xavier Giroud* Ernest Liu† Holger Mueller§

March 2025

Abstract

The vast majority of U.S. inventors work for firms that also have inventors and plants in other tech clusters. Using merged USPTO–U.S. Census Bureau plant-level data, we show that larger tech clusters not only make local inventors more productive but also raise the productivity of inventors and plants in other clusters, which are connected to the focal cluster through their parent firms' networks of innovating plants. Cross-cluster innovation spillovers do not depend on the physical distance between clusters, and plants cite disproportionately more patents from other firms in connected clusters, across large physical distances. To rationalize these findings, and to inform policy, we develop a tractable model of spatial innovation that features both within- and cross-cluster innovation spillovers. Based on our model, we derive a sufficient statistic for the wedge between the social and private returns to innovation in a given location. Taking the model to the data, we rank all U.S. tech clusters according to this wedge. While larger tech clusters exhibit a greater social-private innovation wedge, this is not because of local knowledge spillovers, but because they are well-connected to other clusters through firms' networks of innovating plants. In counterfactual exercises, we show that an increase in the interconnectedness of U.S. tech clusters raises the social-private innovation wedge in (almost) all locations, but especially in tech clusters that are large and well-connected to other clusters.

* Columbia University, NBER, and CEPR. Email: xavier.giroud@gsb.columbia.edu; † Princeton University and NBER. Email: ernestliu@princeton.edu; § New York University, NBER, and CEPR. Email: hmueller@stern.nyu.edu. We thank seminar participants at Stanford, Princeton, Columbia, NYU, Dartmouth, Ohio State, and many other institutions and conferences, as well as numerous colleagues, for helpful comments and suggestions. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1908. (CBDRB-FY24-P1908-R11479).

1 Introduction

Innovative activity is unevenly distributed across space. In the United States, a few large tech clusters, like those in Silicon Valley, Boston, New York, Los Angeles, San Diego, or Seattle, account for a disproportionate share of innovative output.[1] These large tech clusters not only produce more patents, but they also produce more patents per inventor. Importantly, this is not because more productive inventors select into larger tech clusters, but rather because larger clusters make the inventors more productive (Moretti, 2021). While determining the exact sources of this productivity gain is difficult, the literature has pointed to Marshallian agglomeration externalities, in particular, local knowledge spillovers, given that knowledge is a key input into the innovation process.[2] Examples of knowledge spillovers in the context of innovation range from anecdotal evidence, such as the famous Tom Wolfe quote in Saxenian's (1994) comparative study of Silicon Valley vs. Boston's Route 128, to a large empirical literature beginning with Jaffe, Trajtenberg, and Henderson (1993) showing that patent citations are highly localized.[3] Perhaps the most direct and compelling evidence comes from a fascinating recent study by Atkin, Chen, and Popov (2022), who use smartphone geolocation data to identify face-to-face meetings between workers of different firms in Silicon Valley. The study finds that these face-to-face meetings result in significantly higher subsequent mutual patent citations, which is consistent with local knowledge spillovers.

A little known fact about U.S. patenting activity is that 80.7% of inventors work for firms that also have inventors in other tech clusters.[4] If knowledge diffuses locally across inventors of *different* firms, one would expect that the inventors pass on this knowledge to other productive units within their *own* firm, including inventors in other clusters, generating "cross-cluster innovation spillovers." Within the firm, knowledge diffusion does not need to rely on face-to-face interactions but can take place through conventional channels, such as emails, internal memos, or video conferencing. Also, as knowledge is

[1]See, e.g., Audretsch and Feldman (2004), Carlino and Kerr (2015), and Kerr and Robert-Nicoud (2020).

[2]Some view local knowledge spillovers as an integral part of what defines a tech cluster: "*Our definition of tech clusters emphasizes settings with a frontier edge, and many companies seek insights on emerging possibilities, either through first access to codified knowledge or to tacit knowledge that cannot be written down easily*" (Kerr and Robert-Nicoud, 2020, p. 58).

[3]"*Every year there was some place, the Wagon Wheel, Chez Yvonne, Rickey's, the Roundhouse, where members of this esoteric fraternity, the young men and women of the semiconductor industry, would head after work to have a drink and gossip and brag and trade war stories about phase jitters, phantom circuits, bubble memories, pulse trains, bounceless contacts, burst modes, leapfrog tests, p-n junctions, sleeping sickness modes, slow-death episodes, RAMs, NAKs, MOSes, PCMs, PROMs, PROM blowers, PROM blasters, and teramagnitudes, meaning multiples of a million millions*" (Wolfe, 1983, quoted in Saxenian, 1994, p. 33).

[4]Based on the universe of United States Patent and Trademark Office (USPTO) patent data.

non-rival within the firm, it can be used by other productive units at little or no additional cost, creating significant economies of scale and scope.[5]

This paper provides robust evidence of cross-cluster innovation spillovers. Larger tech clusters not only make local inventors more productive, but they also raise the productivity of inventors in other clusters, who are connected to the focal cluster because they work for a firm that has inventors in both clusters. Crucially, firms only benefit from a larger cluster if they have inventors in that cluster. If a firm has a plant in a cluster but no inventors, then a larger cluster does not benefit the firm's local plant, nor does it benefit other plants (or inventors) of the firm. On the other hand, if the firm has a plant with inventors in a cluster, then a larger cluster still benefits other (distant) plants of the firm, by raising plant-level productivity, even if the (receiving) plants do not have inventors. Inventors thus effectively act as "antennas," who process local knowledge and pass it on to other productive units within the firm. Receiving plants, on the other hand, do not necessarily need to have inventors, suggesting that some of the shared knowledge is directly beneficial to the firm beyond raising inventors' productivity.

Our interpretation of inventors as "antennas" is consistent with Cohen and Levinthal's (1989) notion of "absorptive capacity." In an influential paper, the authors posit that "*firms invest in R&D not only to pursue directly new process and product innovation, but also to [...] exploit externally available information*" (p. 593). This externally available information explicitly includes local knowledge spillovers; to benefit from those, firms invest in absorptive capacity:[6]

> "[E]conomists have assumed that technological knowledge which is in the public domain is a public good. Like a radio signal or smoke pollution, its effects are thought to be costlessly realised by all firms located within the neighbourhood of the emission. [...] We suggest that if these costs are relatively small, it is by virtue of the considerable R&D already conducted by the firms in the vicinity of the 'emission'; the firm has already invested in the development of its absorptive capacity in the relevant field" (p. 570).

To explore the role of inventors and plants in fostering cross-cluster innovation spillovers, we merge USPTO patent data with confidential plant-level data from the U.S. Census

[5]Markusen's (1984) model of economies in multi-plant firms is an early example of the idea that knowledge is non-rival within the firm: "*Once an innovation is made, it can be incorporated into any number of additional plants without reducing the marginal product of that innovation in existing plants*" (p. 207).

[6]In Cohen and Levinthal's (1989) model, firms invest in R&D to absorb external knowledge spillovers. In practice, one can think of a firm's "absorptive capacity" as consisting of inventors and other employees involved in the R&D process.

Bureau. Building on (and expanding) the empirical research design in Moretti (2021), we identify cross-cluster spillovers through within-plant variation in the size of "connected clusters"—the number of *other* firms' inventors in the plant's research field in *other* cities where the plant's parent firm also has plants with inventors. That is, we examine if, e.g., inventors and plants in Silicon Valley are more productive if other firms have more inventors in Boston's Route 128, to which the focal (Silicon Valley) inventors and plants are connected through their parent firms' networks of innovating plants. Threats to identification come from omitted factors that may simultaneously affect inventor- or plant-level productivity and the in- or outflow of other firms' inventors in connected clusters. For example, given that the firm's network of innovating plants is non-random, cities in which the firm has innovating plants may be subject to correlated (city-field specific) technology shocks. As a result of these shocks, plant-level productivity may increase, while at the same time clusters to which the plant is connected may experience an inflow of inventors. Our empirical design accounts for these and other omitted factors by including highly granular city $\times$ field $\times$ year fixed effects, besides plant fixed effects. Effectively, we are comparing plants in the same city, research field, and year, which are connected to different clusters, in different cities, because they belong to different parent firms with different spatial networks of innovating plants.

Comparing plants in the same city, research field, and year may not be enough if technology shocks occur at an even more more granular level. To account for this possibility, we propose an instrumental variable (IV) design that isolates variation in the size of a plant's connected clusters that plausibly originates from outside those clusters. The IV design is based on firms that have inventors in multiple locations, including in the plant's connected clusters, but not in the plant's city. The idea is that, when these firms expand (or shrink) their R&D activities in the United States, changes in the number of inventors employed by these firms in *other* cities—where the plant's parent firm has no presence—are predictive of the size of the plant's connected clusters but are unlikely to be systematically correlated with unobserved shocks to the plant's productivity.

In our baseline specification, the elasticity of plant-level total factor productivity (TFP) with respect to the total size of connected clusters is 0.012, and the corresponding elasticity for inventor-level productivity (patents per inventor), aggregated to the plant level, is 0.021. These elasticities are economically meaningful. Suppose we rank plants based on the size of their connected clusters, and consider a plant at the 25th percentile of the corresponding distribution. If we replaced this plant's connected clusters with those of a plant at the 75th percentile, plant-level inventor productivity would increase by 8.2% and plant-level TFP would increase by 4.7%. Accordingly, plants at the top of the (connected-cluster size) distribution—i.e., plants which are either connected to many clusters, to large clusters, or both—experience significant productivity advantages due to cross-cluster innovation spillovers. Our IV estimates are slightly noisier but similar to the OLS estimates (0.014 and 0.023, respectively), suggesting that unobserved shocks to plant-level productivity are not a major source of bias in our empirical specification.

Inventors play a key role in fostering cross-cluster innovation spillovers. If we form "placebo connected clusters" based on cities in which the parent firm has plants but no inventors ("non-innovating plants"), we find no evidence of cross-cluster innovation spillovers. On the other hand, we find that non-innovating plants do benefit from spillovers originating from clusters in which the firm has innovating plants: the corresponding TFP elasticity is 0.006, which is half of the TFP elasticity for innovating plants. Hence, some of the information that is disseminated via cross-cluster spillovers is relevant for productivity more broadly, that is, beyond raising inventor-level productivity.

Plants only benefits from exposure to connected clusters if the parent firm has inventors in these clusters. In the language of Cohen and Levinthal (1989), inventors provide firms with "absorptive capacity" to collect and process local knowledge, and ultimately to pass on this knowledge to other units within the firm. We provide two additional pieces of evidence consistent with knowledge diffusion as a mechanism:

1) Knowledge flows within the firm, across different clusters, can take place trough conventional channels, such as emails, internal memos, or video conferencing. Therefore, if the mechanism is knowledge diffusion within the firm, we would expect cross-cluster innovation spillovers not to decay with physical distance. We confirm this in our data: all our estimates remain stable and significant if we exclude connected clusters within a 100, 250, or 500 mile radius around the plant.

2) As the size of a plant's connected clusters increases, the plant not only cites more patents (by other firms) from those clusters—which could simply be because it innovates more, and therefore cites more patents from everywhere—but the *share* of citations to patents from connected clusters increases. Hence, plants disproportionately cite patents from clusters they are connected to through their firm's innovation network, across large physical distances, which is again consistent with knowledge diffusion.

To rationalize our empirical findings, and to inform policy, we develop a tractable model of spatial innovation that features both within- and cross-cluster innovation spillovers. In the model, fims can invest in innovation to improve plants' productivity. While firms internalize spillovers on other plants of the firm, they do not internalize spillovers on other firms. Cross-cluster spillovers arise for two reasons. First, a plant's innovation investment spills over to other local plants, and from there to other plants of those firms, in other clusters. Second, a plant's innovation investment spills over to other plants of the same firm, in other clusters, and from there to other firms' plants in those clusters. Within- and cross-cluster spillovers interact in a recursive fashion, generating higher-order spillovers that propagate throughout the economy.

We first characterize the government's optimal innovation policy.[7] In the absence of cross-cluster spillovers, the optimal innovation policy involves a uniform subsidy to all innovating plants. However, when there are cross-cluster spillovers, a uniform subsidy is not optimal. Instead, the optimal subsidy is plant-specific and disproportionately targets plants that are well-connected to other clusters. To inform policy as to which locations under-invest the most in innovation, we subsequently derive a model-based sufficient statistic—the "social-private innovation wedge"—which captures the wedge between the social and private returns to innovation in a location. This sufficient statistic can be measured in the data using information on how a given location is connected to other locations through firms' innovation networks in combination with our reduced-form within- and cross-cluster innovation spillover elasticities.

In the final part of the paper, we take our model to the data. We first rank all U.S. tech clusters based on their social-private innovation wedge. While larger clusters exhibit a greater wedge, this is not because of local knowledge spillovers. In fact, if it was only for local knowledge spillovers, the social-private innovation wedge would be *invariant* to cluster size.[8] Instead, what is driving the location ranking is the extent to which clusters are connected to other clusters through firms' networks of innovating plants. Hence, the Silicon Valley cluster has a high social-private innovation wedge not because it generates large local knowledge spillovers, but because it is connected to many other clusters, which are in turn connected to many other clusters, etc. We conclude our quantitative analysis with a counterfactual exercise in which we increase the interconnectedness of U.S. tech clusters. While this raises the social-private innovation wedge in (almost) all clusters—thereby exacerbating the under-investment problem—it especially raises the wedge in clusters that are large and well-connected to other clusters.

Our paper contributes to several strands of literature. First, it contributes to the literature on innovation spillovers. A large literature beginning with Griliches (1979) and Jaffee (1986) shows that firms benefit from other firms' R&D activities. In an important

[7]Spillovers are often viewed as a main justification for innovation policy. As Bloom, Van Reenen, and Williams (2019) note: "*Knowledge spillovers are the central market failure on which economists have focused when justifying government intervention in innovation*" (p. 166).

[8]This invariance echoes a well-known result in urban economics (going back to Glaeser and Gottlieb, 2009) that place-based redistributions are not welfare-enhancing unless the elasticity of productivity to agglomeration differs across locations.

paper, Bloom, Schankerman, and Van Reenen (2013) separate between the (positive) knowledge spillover effect from other firms' R&D and the (negative) business stealing effect.[9] While the R&D spillover literature focuses on spillovers from other firms' R&D, Matray (2021) focuses on spillovers from other firms' *patents*—specifically, how patenting activity by public firms affects the patening activity of local, private firms. Finally, Moretti (2021) focuses on spillovers from other firms' local *inventors.* Building on, and expanding, this empirical research design, we show that these innovation spillovers can propagate throughout the entire economy—leading to more innovative output in distant plants, in other tech clusters—through firms' geographical networks of innovating plants. Furthermore, by calibrating a model of spatial innovation that features both within- and cross-cluster spillovers, we show that cross-cluster innovation spillovers can generate large differences in the wedge between the social and private returns to innovation across different U.S. tech clusters.

Second, our paper contributes to a large literature on the local and firm-level determinants of corporate innovation. Prior studies have examined firms' credit access (Amore, Schneider, and Žaldokas, 2013; Cornaggia et al., 2015; Hombert and Matray, 2017), listing status (Bernstein, 2015), ownership structure (Aghion, Van Reenen, and Zingales, 2013; Li, Liu, and Taylor, 2023), organizational structure (Seru, 2014), and local labor markets (Derrien, Kecskés, and Nguyen, 2023), to name just a few examples. Our paper focuses on the interaction between two determinants: the size (i.e., number of other firms' inventors) of the local tech cluster—from where knowledge "spills over" to the firm's local plant—and the firm's spatial network of innovating plants, through which the knowledge spreads to other plants of the firm, and therefore ultimately to other tech clusters.

Finally, most firm-level studies of regional spillovers examine spillovers *between* firms. By contrast, relatively few studies analyze spillovers across regions through *within-firm* networks (e.g., Cravino and Levchenko, 2017; Giroud and Mueller, 2019; Bena, Dinc, and Erel, 2022; Giroud et al., 2024; Biermann and Huber, 2024). Relative to those studies, our paper shows that within-firm plant-level networks facilitate innovation spillovers across tech clusters. Crucially, plant-level networks alone are not sufficient: plants also need to have inventors, who absorb and process local knowledge and then pass it on to other plants and inventors within the firm, in distant clusters.

The remainder of this paper is organized as follows. Section 2 presents the data,

[9]The typical approach in the R&D spillover literature involves regressing firm outcomes on a weighted average of other firms' R&D capital stocks ("spillover pool"). Arqué-Castells and Spulber (2022) argue that this approach captures not only (genuine) technology spillovers but also (voluntary) technology transfers. Liu and Ma (2024) embed cross-sector R&D spillovers in an endogendous growth model to study the optimal allocation of R&D resources.

analysis sample, and descriptive statistics. Section 3 considers within-cluster innovation spillovers. Section 4 analyzes cross-cluster innovation spillovers; it contains our main results, robustness exercises, results for non-innovating plants, and IV estimates. Section 5 explores potential mechanisms. Section 6 develops a tractable model of spatial innovation and derives the "social-private innovation wedge." Section 7 ranks all U.S. tech clusters based on this wedge and performs counterfactual exercises. Section 8 concludes.

2 Data

Data Sources

We combine data from the United States Patent and Trademark Office (USPTO) patent database with confidential establishment-level data from the U.S. Census Bureau's Longitudinal Business Database (LBD), the Census of Manufactures (CMF), and the Annual Survey of Manufactures (ASM). The patent data include the application date, the date when the patent is granted, the names of inventors and assignees (e.g., firms, universities), other patents cited, research field, and the inventors' home address. The LBD covers all business establishments in the United States with at least one paid employee and contains longitudinal establishment identifiers along with information on employment, payroll, industry, location, and firm affiliation. The CMF is conducted every five years ("Census years") and covers all manufacturing plants in the United States with at least one paid employee. The ASM is conducted annually in all non-Census years and covers a subset of the plants covered by the CMF. Plants with at least 250 employees are included in every ASM year, whereas plants with fewer employees are randomly (re-)sampled every five years. Although the ASM is technically referred to as a "survey," reporting is mandatory, and fines are levied for misreporting. The CMF and ASM contain information about key plant-level variables, such as shipments, assets, material inputs, employment, payroll, capital expenditures, industry, and location.

Sample

We merge the USPTO patent data with the LBD for the years 1976 to 2018. "Year" refers to the year of the patent application, not the year in which the patent is granted. The key challenge is to merge firms in both data sets. For the years 2000 to 2018, we use the bridge created by the U.S. Census Bureau; it is the best available bridge to date.[10] For the years 1976 to 1999, we use the bridge by Kerr and Fu.[11] This bridge is only available for firms in the NBER patent database; we thus extend it to all assignees in the USPTO database following the steps in Section 3 of their paper. We assign patents to research fields using 1-digit Cooperative Patent Classification (CPC) codes. Finally, as in Moretti (2021), we assign patents to Bureau of Economic Analysis (BEA) economic areas ("cities") based on the inventor's (not the assignee's) residential address. BEA economic areas are similar to metropolitan statistical areas (MSAs), albeit some of them are larger. A main advantage of using BEA economic areas, as opposed to MSAs, is that they provide a complete partition of the United States. Figure A.1 in the Online Appendix shows a map of the United States with all 179 BEA economic areas.

We next assign firm inventors to establishments in the LBD. We first aggregate the patent-level data to the inventor-year level. If a patent has multiple inventors, we assign each inventor an equal fraction of the patent. If an inventor has patents in multiple fields in the same year, we use the modal field. We then match inventors to establishments by locating the firm establishment that is nearest to the inventor's residential address. If an inventor has multiple addresses in the same year, we use the modal address. However, inventors may get matched to different establishments in *different* years if they change their home address or, likewise, their firm affiliation. We drop matches if the firm has no establishment in the inventor's city, as it is unclear where the inventor's place of work might be in that case.

In order to be able to compute total factor productivty (TFP) at the establishment level, we link establishments in the LBD to the ASM and CMF (using the common identifier LBDNUM). Thus, our sample consists of manufacturing plants; establishments which are not in the ASM or CMF are being dropped. In some cases, this means that inventor-establishment matches are dropped even if the firm has a manufacturing plant in the same city, because this plant is not the "nearest match." In robustness exercises, we show that our results are similar if, instead of dropping these matches, we assign inventors to the firm's nearest manufacturing plant in the same city. Also, we provide results for *all* inventor-establishment matches in the LBD—i.e., without restricting the sample to manufacturing plants—with the caveat that, in this case, we can only compute inventor productivity (aggregated to the establishment level) but not establishment-level

[10]Dreisigmeyer, David, Nathan Goldschlag, Marina Krylova, Wei Ouyang, and Elisabeth Perlman, 2018, Building a Better Bridge: Improving Patent Assignee-Firm Links, CES Technical Notes, CES-TN-2018-01, U.S. Census Bureau.

[11]Kerr, William, and Shihe Fu, 2008, The Survey of Industrial R&D—Patent Database Link Project, Journal of Technology Transfer 33, 173–186.

TFP. Finally, we aggregate observations to the plant-year level by counting the number of inventors and patents associated with a plant in a given year. If a plant has patents in multiple fields in the same year, we define the plant's field based on its modal field, and we only consider patents in that field.[12] In robustness exercises, we show that our results are similar if we consider all fields in which the plant has patents.

Applying these filters yields a sample of 134,000 plant-year observations. We refer to this sample as "within-cluster spillover" sample. As inventors and patents are always jointly observed in USPTO data, all plant-year observations will have both inventors and patents. In our main analysis sample, we only keep plant-year observations if the firm has at least one other plant with inventors and patents in the same field but in a different city in the given year. This sample has 57,000 plant-year observations; we refer to it as "cross-cluster spillover" sample. Hence, the within-cluster spillover sample includes all plants with matched inventors and patents, while the cross-cluster spillover sample only includes plants if the firm has at least one other plant with matched inventors and patents in the same field but in a different city.

Cluster Size

As in Moretti (2021), a cluster represents a combination of city and research field. At the plant level, a cluster comprises all inventors of *other* firms in the plant's city and research field, i.e., excluding the plant's own inventors as well as any inventors working for the plant's parent firm.[13] In our empirical analysis, we identify innovation spillovers through within-plant variation in cluster size. In this regard, a concern is that cluster size may be mismeasured, as inventors in the USPTO patent data are only observed in years when patents are filed, even though they may also work in the plant's cluster in between patent filings. As we do not know with certainty if this is the case, we choose not to interpolate the data in "missing" (i.e., non-patenting) years. In robustness exercises, we show that our results are similar if we employ a broader notion of cluster size where it is assumed that inventors work in a cluster also in between patent filings.

[12] In our main analysis sample, 86.8% of a plant's patents are in its (modal) field.

[13] We form clusters based on all inventors in the USPTO patent data, not just based on the inventors in our analysis sample. The number of inventors in a city depends on demand and supply factors, including business (demand) and personal (supply) taxes (Moretti and Wilson, 2017), as well as feedback effects: as the number of inventors in a city increases, so do place-based costs, like wages, land prices, and rents, which will in turn affect the number of inventors (Gruber, Moretti, and Wilson, 2023). As long as these factors vary at the city $\times$ year (or even city $\times$ field $\times$ year) level, they are absorbed by our fixed effects.

Descriptive Statistics

Table 1 shows descriptive statistics for plants and firms in our main analysis sample with 57,000 plant-year observations. The typical plant has 699 employees and $275 million in shipments (i.e., sales). Also, it has 4.99 inventors and produces 10.87 patents per year.[14] Finally, the typical plant is "connected" to 3.22 other clusters—other cities in which the parent firm also has plants (in the same field) with inventors—and these other clusters have in total 1,915 inventors (again, in the plant's field) working at other firms. There is substantial variation in the number of connected clusters. Plants in the top decile of the distribution are connected to 11.3 other clusters, and these other clusters have 8,563 inventors of other firms, whereas plants in the bottom decile are only connected to one other cluster, and that cluster has 32 inventors of other firms.

Plants' parent firms have 6,853 employees and $2,524 million in sales. They have 17.51 plants, which are spread across 15.23 counties, 11.77 cities, and 9.16 states. These are *all* plants of those firms, not just plants with inventors ("innovating plants"). At the firm level, the ratio of innovating plants to the total number of plants ("inventor share") is 0.56. This ratio is high for two reasons. First, to be included in our main analysis sample, firms must have at least two innovating plants (in the same field but different cities). Second, smaller firms with only a few plants have a high inventor share by construction. For example, firms with only two plants have an inventor share of 100%, and those with three plants have an inventor share of either 66.6% or 100%. Since all firms receive equal weight, this pushes the average inventor share upward. Alternatively, we could simply divide the total number of innovating plants across all firms in our sample by the total number of plants. Doing so would yield an inventor share of 0.35.

Innovating plants are geographically dispersed within the firm. For the typical firm, the ratio of counties where the firm has innovating plants to the total number of counties where it has plants is 0.55. The corresponding ratio at the city and state level is 0.52 and 0.49, respectively. Hence, firms have innovating plants in about half of the counties, cities, or states in which they have plants. Again, these are averages across all firms, which are high for the reasons discussed above. If we instead divided the total number of counties, cities, or states in which firms have innovating plants by the total number of counties, cities, or states in which they have plants, we would obtain ratios of 0.35, 0.32, and 0.32, respectively. Regardless of which method we use, however, these results strongly suggest that firms do

[14]This is the total number of patents per plant and year, not the number of patents adjusted for co-authorship. That number is 3.36, which implies 0.67 ("fractional") patents per inventor and year. By comparison, in Moretti's (2021) sample of "star inventors"—who are at or above the 90th percentile in terms of patent output—the corresponding number is 1.08 (fractional) patents per inventor and year.

not concentrate their innovative activities in just one or two regions.

3 Within-Cluster Innovation Spillovers

To provide a benchmark, and to illustrate the empirical research design, we first analyze the effect of local cluster size on local plants' innovation and productivity. A larger local cluster means that there are more inventors (at other firms) in the plant's city and field, generating more scope for local knowledge spillovers. We estimate:

$$y_{icft} = \gamma_i + \gamma_{ct} + \gamma_{ft} + \gamma_{cf} + \beta \, Cluster\ Size_{icft} + \varepsilon_{icft}, \qquad (1)$$

where y_{icft} represents either inventor productivity, aggregated to the plant level, or plant-level TFP, i denotes plants, c denotes cities, f denotes research fields, and t denotes years, $Cluster\ Size_{icft}$ is the number of other firms' inventors in the plant's city and field in a given year (in logs), and γ_i, γ_{ct}, γ_{ft}, and γ_{cf} are plant, city $\times$ year, field $\times$ year, and city $\times$ field fixed effects, respectively. Inventor productivity is the ratio (in logs) of the number of patents to the number of inventors at the plant level. This ratio is always non-zero in our sample as patents and inventors are jointly observed in the USPTO patent data and we do not interpolate the data in "missing" (i.e., non-patenting) years. TFP is the estimated residual from a plant-level regression of output on capital, labor, and material inputs (all in logs). To allow for different factor intensities across industries and over time, we estimate the regression separately for each 3-digit SIC code industry and year. Accordingly, TFP can be interpreted as the relative productivity of a plant within a given industry and year. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level.

Our empirical strategy of identifying innovation spillovers through within-unit variation in cluster size builds on Moretti (2021), who focuses on individual inventors as the unit of analysis. In our case, the unit of analysis is a plant. Cluster size varies at the city $\times$ field $\times$ year level, allowing us to include granular city $\times$ year, field $\times$ year, and city $\times$ field fixed effects, besides plant fixed effects. The city $\times$ year fixed effects absorb local shocks (across all fields) that may cause an inflow of inventors in the plant's city while at the same time raising plant-level productivity. Similarly, the field $\times$ year fixed effects absorb technology shocks (across all cities) that may lead to an inflow of inventors in the plant's research field while at the same time raising plant-level productivity. Finally, the city $\times$ field fixed effects absorb time-invariant heterogeneity that may explain why some fields are more productive (and attract more inventors) in certain cities.

Table 2 shows the results. In columns (1) and (2), we replace the city × year, field × year, and city × field fixed effects with field and year fixed effects. (City fixed effects are collinear with the plant fixed effects as plants cannot physically change their location.) In columns (3) and (4), we estimate the full specification from equation (1) with all interacted fixed effects. As is shown, the estimates in columns (1) and (2) are similar to those in columns (3) and (4), suggesting that the estimated elasticities are not very sensitive to shocks at either the city or field level. In column (3), the elasticity of inventor productivity at the plant level with respect to cluster size is 0.076, which is close to the corresponding elasticity of 0.0676 at the individual inventor level in Moretti (2021, column (8), Table 3), despite differences in both sample size and composition resulting from different levels of aggregation and different sample selection criteria. In column (4), the elasticity of plant-level TFP with respect to cluster size is 0.023, suggesting that the same economic forces that make plant inventors more productive also increase overall plant-level productivity.[15] Finally, in column (5), we consider the effect of cluster size on plant-level productivity for plants without inventors ("non-innovating plants").[16] For these plants, "field" is not defined, so we form local clusters based on the *firm's* main field—the modal field based on all firm patents in a given year.[17] As is shown, for non-innovating plants, local cluster size has no significant effect on plant-level TFP. Hence, a plant only benefits from the local presence of other firms' inventors if the plant also has inventors.

4 Cross-Cluster Innovation Spillovers

If local inventors pass on knowledge to other inventors and plants of the firm, including inventors and plants in other clusters, then a larger local cluster should not only affect innovation and productivity at local plants but also at distant plants of the same firm. To examine whether plants of the same firm, in other cities, benefit from a larger local cluster, we estimate the following variant of equation (1):

$$y_{icft} = \gamma_i + \gamma_{cft} + \beta\, Cluster\ Size_{icft} + \varepsilon_{icft}, \tag{2}$$

[15] Since inventor- and plant-level productivity are both outcome variables, this interpretation is consistent with our empirical specification. That said, it is very well possible that plant-level productivity increases, at least in part, *because* the plant innovates more.

[16] Strictly speaking, "non-innovating plants" are de facto "never-innovating plants."

[17] In our within-cluster (cross-cluster) spillover sample, 83.3% (91.1%) of a firm's patents are in its main (i.e., modal) field.

where $Cluster\ Size_{icft}$ is the total size of all clusters in *other* cities where the plant's parent firm has innovating plants (in the plant's field) in a given year (in logs). For simplicity, we refer to those other clusters as "connected clusters." Crucially, connected clusters do not include other cities in which the firm has plants but no inventors. As we have seen previously, non-innovating plants do not benefit from the local presence of other firms' inventors, so we would not expect those plants to pass on local knowledge to other inventors and plants of the firm. (We will confirm this in Table 4 below.)

Cross-cluster innovation spillovers are identified through within-plant variation in the number of other firms' inventors in connected clusters. As we will show below, our results are robust if we focus on connected clusters that are more than 500 miles away from the focal plant. Importantly, within a given city, field, and year, there is ample variation in the size of plants' connected clusters. Accordingly, we can include highly granular city × field × year fixed effects—denoted by γ_{cft} in equation (2)—allowing us to separate cross-cluster innovation spillovers from productivity shocks that are common to the locations of the firm's innovating plants.[18] For example, given that the firm's network of innovating plants is non-random, cities in which the firm has innovating plants may be subject to correlated city-field specific technology shocks. As a result, the focal plant's productivity may increase, while at the same time other cities to which the plant is connected may experience an inflow of inventors. Effectively, we are comparing plants in the same city, research field, and year, which are connected to different clusters because they belong to different parent firms with different spatial networks of innovating plants.

Main Results

Table 3 provides strong and robust evidence of cross-cluster innovation spillovers. As we did in Table 2, we successively add more granular fixed effects, starting with field and year fixed effects (columns (1) and (2)), then including city × field, city × year, and field × year fixed effects (columns (3) and (4)), and finally including city × field × year fixed effects (columns (5) and (6)), besides plant fixed effects. The latter is our tightest specification and will be our baseline specification going forward. As can be seen, the point estimates are stable and significant across different fixed-effect specifications, suggesting that, similar to what we found in the within-cluster analysis, the estimated elasticities are not very sensitive to unobserved shocks at the city or field (or even city × field) level.

In columns (5) and (6), the elasticities of plant-level inventor productivity and

[18]In robustness exercises, we show that our results are similar if we use (even more granular) city × class × year fixed effects, where technology classes are defined at the 3-digit CPC code level.

plant-level TFP with respect to the total size of connected clusters are 0.021 and 0.012, respectively. These elasticities are between one third (inventor productivity) and one half (TFP) of the corresponding within-cluster elasticities in columns (3) and (4) of Table 2, suggesting that some, but not all, of the local information is beneficial to other inventors and plants of the firm. (Alternatively, some information may be "soft" and thus difficult to pass on to other firm units.) To interpret the magnitudes, suppose we rank all plants based on their value of $ClusterSize_{icft}$, and consider a plant at the 25th percentile of the corresponding distribution. If we replaced this plant's connected clusters with those of a plant at the 75th percentile, plant-level inventor productivity would increase by 8.2% and plant-level TFP would increase by 4.7%. Hence, plants at the top of the distribution—i.e., plants which are either connected to many clusters, to large clusters, or both—experience significant productivity advantages due to cross-cluster innovation spillovers.

Non-Innovating Plants

As we previously showed in column (5) of Table 2, non-innovating plants do not benefit from the local presence of other firms' inventors, so we would not expect these plants to pass on local knowledge to other inventors and plants within the firm. To verify this conjecture, we link (innovating) plants to "placebo clusters" where the parent firm has plants but no inventors. As this placebo test only requires firms to have at least one innovating plant—as opposed to our main analysis sample, which requires firms to have innovating plants in at least two cities—the number of observations is larger than in our main sample. Table 4 shows the results. As columns (1) and (2) show, a larger cluster in cities where the parent firm has non-innovating plants—even if the cluster is in the focal plant's research field—has no significant effect on either plant-level inventor productivity or plant-level TFP. We may thus conclude that cross-cluster innovation spillovers do not simply originate from any cluster in which the parent firm has plants but only from those clusters where it has innovating plants.

Do non-innovating plants benefit from spillovers originating from clusters in which the firm has innovating plants? A priori, the answer is unclear. If cross-cluster spillovers only raise plant-level productivity *because* they spur innovation, then non-innovating plants should not benefit from such spillovers. On the other hand, if cross-cluster spillovers disseminate information that is relevant for productivity more broadly, then even non-innovating plants may benefit. We explore this hypothesis in column (3). As "field" is not defined for plants without inventors, we use again the firm's main field to form connected clusters, analogous to what we did in column (5) of Table 2. As is shown, non-innovating plants do benefit from cross-cluster spillovers, albeit less than innovating plants; the elasticity of plant-level TFP with respect to the total size of connected clusters is one half of the corresponding TFP elasticity in column (6) of Table 3. Hence, some of the information that is disseminated through cross-cluster spillovers is relevant for productivity more broadly, that is, beyond raising inventor-level productivity.

Robustness

Table A.1 of the Online Appendix contains miscellaneous robustness tests. In our main sample, we drop inventor-establishment matches if the ("nearest-match") establishment is not a manufacturing plant, even if the firm has a manufacturing plant in the same city. In columns (1) and (2), instead of dropping these matches, we assign inventors to the firm's nearest manufacturing plant in the same city. As can be seen, the results are similar to our baseline results. Second, in the USPTO patent data, inventors are only observed in years when patents are filed. In our main sample, this implies that cluster size may be mismeasured if inventors work in a cluster also in between patent filings. In columns (3) and (4), we employ a broader notion of cluster size which assumes that inventors work in the cluster also in between filing years (provided they do not file patents elsewhere during that time). Again, the results are similar to our baseline results.

Table A.2 of the Online Appendix contains robustness exercises pertaining to patent fields and industry coverage. In our main sample, the plant's field is based on the modal field across all its patents in a given year. This has two implications: i) we only consider patents in that (modal) field, and ii) for consistency, the plant's connected clusters only consist of other firms' inventors in that field. In columns (1) and (2), we consider all fields in which the plant has patents, and connected clusters consist of other firms' inventors in all those fields. As is shown, the results are similar to our baseline results. In columns (3) and 4), we replace the city × field × year fixed effects with (even more granular) city × class × year fixed effects, where technology classes are defined at the 3-digit CPC code level. Once again, the results are similar to our baseline results.

Our final robustness exercise pertains to industry coverage. Our main sample is restricted to manufacturing establishments in the ASM and CMF, which allows us to compute plant-level TFP. We are interested in plant-level TFP for two reasons. First, for innovating plants, it provides us with an additional measure of productivity besides inventor productivity. Second, for non-innovating plants, it provides us with *a* measure of productivity, as inventor productivity is not available for those plants. That said, while many innovating firms are in manufacturing industries, others are not.[19] In column (5), we drop the sample requirement that LBD establishments must have a link to the ASM and CMF. Accordingly, we consider all industries, not just manufacturing industries. All other data filters and sample selection criteria remain the same. Naturally, this means we can only examine the effects on inventor productivity (aggregated to the establishment level) but not on TFP. As can be seen, the result is similar to our baseline result in column (5) of Table 3, despite a much broader industry coverage and a much larger sample size.

Instrumental Variable Estimation

Our empirical specification accounts for unobserved shocks at the granular city × field × year level (or even city × class × year level; see Table A.2) that may cause an inflow of other firms' inventors in connected clusters while at the same time raising the plant's productivity. However, equation (2) may lead to biased estimates if the shocks occur at an even more granular level. To account for this possibility, we propose an instrumental variable (IV) design that isolates variation in the size of a plant's connected clusters that plausibly originates from outside those clusters. The idea is that, when firms that have inventors in the plant's connected clusters (but not in the plant's city) expand (or shrink) their R&D activities in the United States, changes in the number of inventors employed by those firms in *other* (i.e., "third") cities—where the plant's parent firm has no physical presence—will be predictive of the size of the plant's connected clusters. The identifying assumption is that factors that drive the expansion of those other firms in other cities are not systematically correlated with unobserved shocks to the (focal) plant's productivity, conditional on fixed effects.[20]

To illustrate, consider a plant in Silicon Valley that is connected to the San Diego and Seattle clusters through its parent firm's network of innovating plants. Other firms which have inventors in the San Diego and Seattle clusters also have inventors in the Boston and New York clusters (where the Silicon Valley plant's parent firm has no presence). The first-stage regression exploits changes in the number of those other firms' inventors in Boston and New York to predict (within-plant) changes in the size of the San Diego and Seattle clusters; instrument exogeneity requires that the growth in the number of other

[19]Manufacturing industries with a particularly high share of U.S. patenting output include, e.g., pharmaceuticals and medicines, medical equipment, aerospace products, automobiles, communications equipment, and semiconductors. Manufacturing industries account for 66.3% of all U.S. patents issued in 2013 (Table 60, Business R&D and Innovation: 2013, Detailed Statistical Tables NSF 16-313, National Science Foundation, National Center for Science and Engineering Statistics, 2016).

[20]Moretti (2021) uses a similar IV strategy to instrument for local cluster size.

firms' inventors in Boston and New York is not systematically correlated with unobserved shocks to the productivity of the Silicon Valley plant.

Table 5 presents the results. As column (1) shows, the number of other firms' inventors in cities where the plant's parent firm has no physical presence is highly predictive of the size of the plant's connected clusters. The instrument is both relevant and strong; the first-stage F-statistic is 70.6. Columns (2) and (3) present 2SLS estimates for plant-level inventor productivity and TFP, respectively. As is shown, the estimates are similar to the corresponding OLS estimates (naturally, the IV estimates are noisier), suggesting that unobserved shocks to plant-level productivity, conditional on fixed effects, are unlikely to be a major source of bias in our empirical specification.

Number of Plant Inventors

In our theoretical framework, we assume that the effect of cluster size on plant-level productivity scales up with the plant's "innovation investment" (e.g., inventors, R&D). To evaluate this assumption empirically, we interact $ClusterSize_{icft}$ with the number of inventors (in logs) at the plant level. The results are shown in Table A.3 of the Online Appendix. As our model features both within- and cross-cluster innovation spillovers, we show empirical results for both. In both cases, we consider the number of inventors at the focal plant (columns (1) to (4)); in the case of cross-cluster spillovers, we additionally consider the number of inventors at the firm's (innovating) plants in connected clusters (columns (5) and (6)). There are two main takeaways from Table A.3. First, having more inventors at a plant is associated with higher plant-level inventor productivity and higher plant-level TFP, which is consistent with peer effects. (We are obviously careful not to make causal statements here.) Second, and importantly, the interaction term between $ClusterSize_{icft}$ and the number of plant inventors is positive, suggesting that the two are complements, which is consistent with our modeling assumption in Section 6.

Aggregation to the Cluster Level

Our plant-level specification makes comparisons between plants in the same cluster based on their differential exposures to other clusters. In Table A.4 of the Online Appendix, we make comparisons between clusters by aggregating plant-year observations to the cluster-year level. Inventor productivity at the cluster level is the ratio (in logs) of the total number of patents to the total number of inventors based on all plants in the cluster. TFP at the cluster level is the employment-weighted average value of TFP across all plants in the cluster. To compute the total size of a cluster's "connected clusters," we either add up (columns (1) and (2)) or compute the average value of (columns (3) and (4)) $ClusterSize_{icft}$ across all plants in the cluster.[21] As can be seen, the estimates are similar to the corresponding estimates at the plant level.

5 Mechanism

Plants only benefit from exposure to other clusters if the parent firm has inventors in those clusters. In the language of Cohen and Levinthal (1989), inventors provide firms with "absorptive capacity" to collect and process local knowledge, and ultimately to pass on this knowledge to other units within the firm. We now present two additional pieces of evidence suggesting that knowledge diffusion is the driving force behind cross-cluster innovation spillovers, based on geographical proximity and backward patent citations.

Geographical Distance

Knowledge diffusion within the firm need not rely on face-to-face interactions but can take place trough "conventional" channels, such as emails, internal memos, or video conferencing. Accordingly, if the driving force behind cross-cluster innovation spillovers is knowledge diffusion within the firm, we would expect the spillovers not to decay with geographical distance. To test this hypothesis, we exclude all connected clusters within a 100, 250, or 500 mile radius around the location of the focal plant. As some plants have all their connected clusters within these radii, the sample gradually shrinks. As is shown in Table 6, all our estimates remain stable if we exclude nearby connected clusters.

Patent Citations

Patent citations are often viewed as evidence of knowledge flows. Hence, if the driving force behind cross-cluster innovation spillovers is knowledge diffusion, we would expect plants to cite more patents from connected clusters as the size of those clusters increases. However, citation counts alone are not informative. This is because, as we have previously shown, an increase in the size of a plant's connected clusters raises the plant's inventor productivity and therefore its patent output.[22] But if the plant produces more patents, it

[21]In columns (1) and (2), if n plants in a cluster are connected to the same distant cluster, this distant cluster appears n times in the summation, thus receiving greater weight. In columns (3) and (4), averages are simple averages, but the results are virtually identical if we use employment-weighted averages.

[22]The results for patent output mirror those for inventor productivity; see Table A.5 of the Online Appendix.

will also cite more patents. Even if the plant were to cite patents equally from all clusters, this would mean it cites more patents from connected clusters, simply because it cites more patents in general. For this reason, we focus not on patent citation counts but on patent citation *shares*. Specifically, we examine if an increase in the size of a plant's connected clusters affects the share of the plant's citations to patents from those clusters.

In Table 7, the dependent variable is the share of the plant's citations to other firms' patents from connected clusters relative to all its citations to other firms' patents (in the plant's field). In columns (1) to (3), we successively add more granular fixed effects. As is shown, the estimates are positive and stable across all specifications. Hence, plants cite disproportionately more patents from connected clusters, across large physical distances, as the size of those clusters increases. In column (4), we drop citations to patents from the same year to account for the contemporaneous growth in patents from connected clusters. As is shown, our results are virtually unchanged.[23] Finally, in column (5), we consider citations to patents from "placebo clusters" in which the firm has plants but no inventors, similar to what we did in columns (1) and (2) of Table 4. As is shown, an increase in the size of those clusters has no significant effect on their share of the plant's patent citations.

6 Theoretical Framework

To rationalize our empirical findings, and to inform policy, we develop a theoretical framework in which firms make innovation investments in the presence of within- and cross-cluster innovation spillovers. A key objective of our model is the derivation of a sufficient statistic that captures the wedge between the social and private returns to innovation in a location. In Section 7, we will take our model to the data and use this wedge to show which U.S. tech clusters under-invest the most in innovation, and also to perform counterfactual exercises where we increase the interconnectedness of tech clusters.

6.1 Setup

There are N locations. Each location has a representative consumer and a continuum of plants producing differentiated varieties that are sold to all locations subject to iceberg trade costs. Each plant belongs to a firm, and each firm owns a set of plants across different locations. Plants can invest in innovation to improve their productivity. Our model features both within-location innovation spillovers—a plant may receive knowledge spillovers from other plants in the same location—and cross-location spillovers—a plant

[23] Our results remain similar if we drop citations to patents from the last two, five, or ten years.

may receive knowledge spillovers from other plants of the same firm, in other locations. Within- and cross-location spillovers may interact, generating higher-order spillovers: knowledge may spill over to other plants of the same firm, in other locations, from there it may spill over to other firms' plants in those locations, and so on.

Preferences. Each location n has a representative consumer with preferences:

$$U_n = \ln Y_n - L_n, \qquad Y_n \equiv \left[\sum_i \int y_{ni}(\nu)^{\frac{\sigma-1}{\sigma}} d\nu\right]^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

where Y_n is a consumption aggregator, $y_{ni}(\nu)$ is the quantity of variety ν produced in location i that is consumed in location n, L_n is the labor supplied in location n, and $\sigma > 1$ is the elasticity of substitution.

The consumer's budget constraint is:

$$\sum_i \int p_{ni}(\nu)\, y_{ni}(\nu)\, d\nu = w_n L_n + \Pi_n - T_n, \tag{4}$$

where $p_{ni}(\nu)$ is the price of variety ν produced in location i that is sold to location n, w_n is the wage rate, which we normalize to one, Π_n are firm profits rebated to the consumer in location n, and T_n is a lump-sum tax.

The solution to the consumer's problem of maximizing (3) subject to (4) yields a demand curve for each variety:

$$p_{ni}(\nu)\, y_{ni}(\nu) = \frac{p_{ni}(\nu)^{1-\sigma}}{\sum_i \int_\nu p_{ni}(\nu)^{1-\sigma} d\nu}, \tag{5}$$

which implies total expenditures by consumer n across all varieties sum up to one:

$$\sum_i \int p_{ni}(\nu)\, y_{ni}(\nu)\, d\nu = 1. \tag{6}$$

Remark. We normalize the population of consumers in each location to one for expositional convenience. As is shown in Online Appendix C.1, our analysis readily extends to a setting with heterogeneous location size.

Production and Trade. Plants are monopolistically competitive and feature constant-returns-to-scale production using labor. Given marginal cost $\kappa_i(\nu)$ in labor units for a plant in location i producing variety ν, the price of its product that is sold to location n is:

$$p_{ni}(\nu) = \frac{\sigma}{\sigma-1}\tau_{ni}\kappa_i(\nu), \tag{7}$$

where $\frac{\sigma}{\sigma-1}$ is the markup and $\tau_{ni} \geq 1$ is a bilateral iceberg trade cost for goods produced in location i that are sold to location n. The plant's variable profit from selling goods to location n is:

$$\pi_{ni}(\nu) = \frac{1}{\sigma}p_{ni}(\nu)\,y_{ni}(\nu). \tag{8}$$

Substituting (5) and (7) in (8) and summing across all locations n which the plant sells to, the plant's total variable profit is:

$$\pi_i(\nu) = b_i\kappa_i(\nu)^{1-\sigma},$$

where b_i is a (location-specific) "market access" demand shifter for plants in location i:

$$b_i \equiv \sum_n \frac{1}{\sigma}\frac{\tau_{ni}^{1-\sigma}}{\sum_m \int_\nu (\kappa_m(\nu)\,\tau_{nm})^{1-\sigma}\,d\nu}. \tag{9}$$

Location i's market access is decreasing in the trade cost of selling goods to other locations, τ_{ni}, and increasing in the production and trade costs of producers from other locations.

Firm Networks, Innovation Investments, and Knowledge Spillovers. Each plant belongs to a firm, and each firm owns a set of plants across different locations. We refer to a plant's productivity as its *knowledge*, which is inversely related to marginal cost. A plant can invest in innovation to improve its knowledge and thus reduce its marginal cost. Motivated by our reduced-form evidence, we consider two types of knowledge spillovers. First, a plant benefits from the knowledge of other local plants. Second, a plant benefits from the knowledge of other plants of the same firm, in other locations.

In our baseline model, we assume no heterogeneity across plants besides their locations and the firms they belong to. We can therefore use the set of plant locations J to denote firms with plants in locations $n \in J$. Likewise, we use nJ to denote the plant in location $n \in J$ that belongs to firm J. Also, we abstract from the extensive margin of innovation and focus on the intensive margin: we assume all plants innovate to some degree in equilibrium. In Online Appendix C.2, we extend our setting by adding heterogeneous productivities and endogenous innovation decisions along the extensive margin, so that only a subset of the plants innovates in equilibrium, and we derive our main sufficient statistic—the social-private innovation wedge—in that extended setting.

Let k_{nJ} denote the knowledge of plant nJ. We specify:

$$k_{nJ} \equiv \prod_{i \in J} \left(h_{iJ}^{\alpha} K_i^{\delta} \right)^{\omega_{ni}}, \tag{10}$$

where h_{iJ} denotes plant iJ's innovation investment and K_i is an aggregator of the knowledge of all plants in location i:

$$K_i \equiv \sum_{J \ni i} N_J k_{iJ}, \tag{11}$$

where N_J is the measure of firms with plants in location set J.

Equations (10) and (11) have five key implications. First, plants directly benefit from their own innovation investment as well as from the innovation investments of other plants of the same firm. Second, plants receive local knowledge spillovers, which are increasing in their own innovation investment. Third, plants receive knowledge spillovers from other locations, which are increasing in their parent firm's innovation investments in those (other) locations. The second and third implications are consistent with the evidence presented in Table A.3 of the Online Appendix. Fourth, the knowledge spillovers which plants receive, both locally and from other locations, are increasing in the innovation investments of other firms. Fifth, the recursive formulation in (10) and (11) implies an infinite loop of knowledge spillovers, where knowledge spills over to a firm's local plant, from there it spills over to other plants of the same firm, in other locations, from there it spills over to other firms' plants in those locations, and so on.

A key object in our model is the strength of knowledge spillovers across locations governed by the elasticities ω_{ni}. Motivated by our reduced-form evidence that spillovers across clusters are weaker than within-cluster spillovers but do not decay with distance, we specify:

$$\omega_{ni} = \begin{cases} 1 & \text{if } n = i, \\ \omega < 1 & \text{otherwise.} \end{cases}$$

A plant's marginal cost is inversely related to its knowledge: $\kappa_{nJ}^{1-\sigma} = k_{nJ}$. We can thus write the variable profit of plant nJ as:

$$\pi_{nJ} = b_n \prod_{i \in J} \left(h_{iJ}^{\alpha} K_i^{\delta} \right)^{\omega_{ni}}, \tag{12}$$

where b_n is the market access in location n, as defined in (9), which can be re-written as:

$$b_i \equiv \sum_n \frac{1}{\sigma} \frac{\tau_{ni}^{1-\sigma}}{\sum_m \sum_{J \ni m} N_J k_{mJ} \tau_{nm}^{1-\sigma}}. \tag{13}$$

Given the demand elasticity σ (c.f. equation 5), plant-level revenues are $r_{nJ} = \sigma \pi_{nJ}$.

The firm internalizes the impact of investing in innovation at any of its plants on the knowledge of any of its other plants and makes investment decisions accordingly to maximize the total profits of all of its plants. However, the firm does not internalize the impact of its innovation investment on location-wide knowledge K_n and thus on the productivities of other firms' plants. Formally, firm J's problem is to choose innovation investments $\{h_{nJ}\}$ across all of its plants to maximize the total profits of these plants, net of the cost of innovation investments:

$$\hat{\pi}_J \equiv \max_{\{h_{nJ}\}} \sum_n \left[b_n \prod_{i \in J} \left(h_{iJ}^{\alpha} K_i^{\delta} \right)^{\omega_{ni}} - \frac{c_n h_{nJ}}{t_{nJ}} \right]. \tag{14}$$

The first term inside the bracket captures plant nJ's variable profits from selling its products, c_n is the marginal cost of innovation investments in labor units, and t_{nJ} is a proportional subsidy to plant nJ's innovation which the firm takes as given.

We assume that all profits are rebated back to consumers, satisfying:

$$\sum_n \Pi_n = \sum_J N_J \hat{\pi}_J. \tag{15}$$

Government Budget. Innovation subsidies are provided by the government, which must balance its budget by levying lump-sum taxes:

$$\sum_n \left\{ \sum_{J \ni n} N_J \left(\frac{c_n h_{nJ}}{t_{nJ}} - c_n h_{nJ} \right) + T_n \right\} = 0. \tag{16}$$

Market Clearing Conditions. Each plant's output must equal the sum of quantities delivered to consumers in all locations. Because plants produce linearly from labor, the labor market clearing condition is:

$$\sum_{iJ} k_{iJ}^{\frac{1}{1-\sigma}} \sum_n y_{niJ} t_{ni} + \sum_{nJ} c_n h_{nJ} = L, \tag{17}$$

where (with a slight abuse of notation) we use y_{niJ} to denote the consumption in location n of the variety produced by plant iJ.

Definition 1. Given innovation subsidies $\{t_{nJ}\}$, an equilibrium is the collection of innovation investments $\{h_{iJ}\}$ that solves the firm's problem (14), where plant-level knowledge k_{nJ} and location-wide knowledge K_i are recursively defined by (10) and (11), the market access b_n satisfies (13), the price of plant iJ's output that is sold in location n is $\frac{\sigma}{\sigma-1}\tau_{ni}k_{iJ}^{\frac{1}{1-\sigma}}$, the consumer in each location chooses consumption and labor supply to maximize (3) subject to (4), lump-sum taxes are levied by the government to balance its budget according to (16), profits are rebated to consumers according to (15), and the labor market clears according to (17).

Definition 2. An equilibrium without government subsidies (i.e., $t_{nJ} \equiv 1\ \forall nJ$) is referred to as *Laissez-faire* equilibrium.

Lemma 1. *Let* $\mathcal{J} \equiv \max |J|$ *denote the maximum number of plants belonging to a firm. A sufficient condition for equilibrium uniqueness is* $(\alpha + \delta)\,(1 + (\mathcal{J} - 1)\,\omega) < 1$.

Proof. See Online Appendix B.1. □

Social Welfare. Utilitarian social welfare is the sum of consumer utilities across locations:

$$\mathcal{W} \equiv \sum_n U_n.$$

Substituting (6) in (4), summing across locations, and substituting out Π_n and T_n using (15) and (16), we can write social welfare as a function of the innovation investments $\{h_{nJ}\}$:

$$\mathcal{W} \equiv \sum_n \left(\ln Y_n - N_J \sum_{J \ni n} c_{nJ} h_{nJ} \right) - \frac{\sigma - 1}{\sigma} N. \tag{18}$$

Equation (18) implies that social welfare is equal to the value of the aggregate consumption bundle net of innovation costs minus a constant. To facilitate the interpretation of our welfare analysis, it is useful to define an alternative social welfare function $\hat{\mathcal{W}}$ that only depends on firms' variable profits and innovation costs:

$$\hat{\mathcal{W}} \equiv \sum_n \left(\frac{\sigma}{\sigma - 1} \sum_{J \ni n} N_J b_n \prod_{i \in J} \left(h_{iJ}^{\alpha} K_i^{\delta} \right)^{\omega_{ni}} - \sum_{J \ni n} N_J c_{nJ} h_{nJ} \right). \tag{19}$$

Using the definition of variable profits in equation (12), we can write $\hat{\mathcal{W}}$ conveniently as:

$$\hat{\mathcal{W}} \equiv \sum_n \sum_{J \ni n} N_J \left(\frac{\sigma}{\sigma - 1} \pi_{nJ} - c_{nJ} h_{nJ} \right).$$

Lemma 2. $\frac{d\mathcal{W}}{dh_{nJ}} = \frac{d\hat{\mathcal{W}}}{dh_{nJ}}|_{holding\ \{b_n\}\ constant}$ *for all* nJ.

Proof. See Online Appendix B.2. □

Lemma 2 implies that, in equilibrium, the marginal impact of innovation investments on social welfare is proportional to the impact on the sum of variable profits ($\frac{d \sum_n \ln Y_n}{dh_{kJ'}} = \frac{\sigma}{\sigma-1} \frac{d \sum_n \sum_J N_J \pi_{nJ}}{dh_{kJ'}}$), taking the demand shifters $\{b_n\}$ in each location as given. This allows us to examine the impact of plant-level innovation on social welfare by simply analyzing the impact on firms' profits. Juxtaposing the welfare function $\hat{\mathcal{W}}$ with the objective function of the firm's problem (14), we can discuss the sources of inefficiencies that arise in a Laissez-faire equilibrium.

6.2 Equilibrium

Sources of Inefficiencies. The Laissez-faire equilibrium features three sources of inefficiencies. First, firms only internalize the impact of their innovation investments on profits and ignore the impact on consumer surplus. This inefficiency is captured by $\frac{\sigma}{\sigma-1}$ in the social welfare function (19) and, by Lemma 2, can be corrected through a uniform innovation subsidy across all plants.

Second, firms only internalize the impact of their innovation investments on their own profits and ignore the impact on other firms' profits. This second inefficiency manifests itself in two ways. First, a plant's innovation investment generates local knowledge spillovers on other firms' local plants. Second, a plant's innovation investment generates knowledge spillovers on other firms' plants in other locations, either through those firms' plant-level networks or the plant's parent firm's network. This second inefficiency can be corrected through innovation policies that subsidize plants or firms heterogeneously depending on their location and "connectedness" to other locations.

Finally, due to markups, firms under-produce given their marginal costs. This inefficiency is standard in monopolistic settings and can be corrected through a uniform production subsidy across all plants. In our analysis, this inefficiency plays no role. Our equilibrium definition takes markups as given, and we do not seek to correct for them. Instead, we focus on innovation subsidies and their impact on social welfare.

Social and Private Value of Innovation. To firm J, the private value of investing in innovation in plant iJ, defined as the semi-elasticity of firm-wide profits with respect to plant-level innovation h_{iJ}, is:

$$\beta_{iJ} \equiv \frac{\partial \sum_{n \in J} \pi_{nJ}}{\partial \ln h_{iJ}} = \alpha \sum_{n \in J} \pi_{nJ} \omega_{ni} \mathbf{1}_{ni}^{J}, \tag{20}$$

where $\{n \in J\}$ is the set of locations in which firm J has plants, and $\mathbf{1}_{ni}^{J}$ is an indicator for firm J having plants in both locations n and i. In equilibrium, by the first-order condition of the firm's problem (14), the private value of innovation must equal the firm's total private expenditure on innovation:

$$\beta_{iJ} = \frac{c_i h_{iJ}}{t_{iJ}} \quad \text{in equilibrium.} \tag{21}$$

The social value of investing in innovation in plant iJ, defined as the semi-elasticity of social welfare with respect to plant-level innovation, is:

$$\begin{aligned} \gamma_{iJ} &\equiv N_J^{-1} \frac{d\hat{\mathcal{W}}}{d \ln h_{iJ}}\Big|_{\{b_n\}} \\ &= \frac{\sigma}{\sigma-1} N_J^{-1} \left(N_J \frac{\partial \sum_{n \in J} \pi_{nJ}}{\partial \ln h_{iJ}} + \sum_m \frac{\partial \sum_{J'} N_{J'} \sum_{n \in J'} \pi_{nJ'}}{\partial \ln K_m} \frac{d \ln K_m}{d \ln h_{iJ}} \right) \\ &= \frac{\sigma}{\sigma-1} N_J^{-1} \left(\alpha \sum_{n \in J} N_J \pi_{nJ} \omega_{ni} \mathbf{1}_{ni}^{J} + \delta \sum_{J'} N_{J'} \sum_{n \in J'} \sum_m \pi_{nJ'} \omega_{nm} \mathbf{1}_{nm}^{J'} \frac{d \ln K_m}{d \ln h_{iJ}} \right). \end{aligned} \tag{22}$$

The first term inside the parenthesis is proportional to the private value of innovation; it captures the impact of plant-level innovation h_{iJ} on the profit of firm J (multiplied by $\frac{\sigma}{\sigma-1}$ to account for the impact on consumer surplus). The second term inside the parenthesis is new. It is proportional to the total indirect effect of plant-level innovation on the profits of all firms through its impact on location-wide knowledge $\{K_m\}$. A social planner seeking to maximize social welfare will choose innovation investments such that the social value coincides with the total social cost of innovation:

$$\gamma_{iJ} = h_{iJ} c_i. \tag{23}$$

The Impact of Plant-Level Innovation on Location-Wide Knowledge. We now derive the impact of plant-level innovation h_{iJ} on location-wide knowledge K_n, both in the plant's own location and in other locations. To this end, we substitute (10) in (11) and re-write location-wide knowledge K_n as:

$$K_n \equiv \sum_{J \ni n} N_J k_{nJ} = \sum_{J \ni n} N_J \prod_{i \in J} \left(h_{iJ}^{\alpha} K_i^{\delta}\right)^{\omega_{ni}}.$$

Totally differentiating and substituting using $r_{nJ} = \sigma b_n \prod_{i \in J} \left(h_{iJ}^{\alpha} K_i^{\delta}\right)^{\omega_{ni}}$ we obtain:

$$\frac{d \ln K_n}{d \ln h_{iJ}} \quad = \quad S_{nJ} \alpha \omega_{ni} \mathbf{1}_{ni}^{J} + \sum_{J' \ni n} S_{nJ'} \sum_{q \in J'} \delta \omega_{nq} \frac{d \ln K_q}{d \ln h_{iJ}}, \tag{24}$$

where $S_{nJ} \equiv \frac{N_J r_{nJ}}{\sum_{J' \ni n} N_{J'} r_{nJ'}}$ denotes plant nJ's sales as a share of total sales by all plants in location n.

Equation (24) decomposes the effect of plant-level innovation h_{iJ} on location-wide knowledge K_n into direct and indirect network effects. A plant's innovation investment improves both its own knowledge and the knowledge of other plants of the same firm, which contributes to location-wide knowledge in the plant's own location and the other plants' locations, including location n. This direct spillover effect is captured by the first term on the right-hand side. An increase in location-wide knowledge in any of these locations in turn improves the knowledge of any plants in these locations, which in turn improves the location-wide knowledge in any of the locations these plants are connected to through their firms' plant-level networks, including location n. The second term on the right-hand side captures this higher-order (indirect) spillover effect. We could further decompose and expand $\frac{d \ln K_q}{d \ln h_{iJ}}$ into direct and indirect spillover effects. Instead, we provide a succinct summary of the infinite rounds of network spillover effects in matrix notation below. For now, we merely note that plant iJ's innovation investment has a larger impact on location-wide knowledge K_n if firm J's plant in location n accounts for a large share of local sales, or if plant iJ has a large impact on location-wide knowledge in other locations q with many plants there being connected to large plants in location n.

Optimal Innovation Subsidy. Comparing the equilibrium and social optimum first-order conditions, (21) and (23), and recognizing that plant-level revenues are proportional to profits ($r_{nJ} = \sigma \pi_{nJ}$), we can now characterize the optimal subsidy.

Proposition 1. *The optimal innovation innovation subsidy* $\left\{\arg\max_{\left\{t_{iJ}\right\}} \hat{\mathcal{W}}\right\}$ *is* *plant-specific and satisfies*

$$t_{iJ} = \frac{\sigma}{\sigma-1}\frac{\left[\alpha N_J \sum_{n\in J} r_{nJ}\omega_{ni}\mathbf{1}^J_{ni} + \delta \sum_{J'} N_{J'} \sum_{n\in J'} \sum_m r_{nJ'}\omega_{nm}\mathbf{1}^{J'}_{nm}\frac{d\ln K_m}{d\ln h_{iJ}}\right]}{\alpha N_J \sum_{n\in J} r_{nJ}\omega_{ni}\mathbf{1}^J_{ni}}. \tag{25}$$

Proposition 1 has two important implications. First, the optimal innovation subsidy is generally not uniform, neither across all plants nor across all plants in a given location. Two plants in the same location may receive different optimal subsidies if they are differentially connected to other locations. Ceteris paribus, a plant should receive a higher subsidy if its innovation investment leads to a larger increase in location-wide knowledge in locations which are disproportionately connected to other locations with high-revenue plants. There are special cases when the optimal subsidy is uniform: when cross-location spillovers are zero ($\omega_{nm} = 0$ for all $n \neq m$) or when all firms are single-location firms ($\mathbf{1}^J_{nm} = 0$ for all $n \neq m$). In those special cases, the optimal subsidy (25) simplifies intuitively to:

$$t_{iJ} = \frac{\sigma}{\sigma-1}\frac{1}{1-\delta} \text{ for all } iJ, \tag{26}$$

where $\frac{\sigma}{\sigma-1}$ corrects for the fact that plants under-invest in innovation because their profits do not capture the full social surplus, and $\frac{1}{1-\delta}$ corrects for the fact that plants under-invest because they ignore knowledge spillovers on other plants through their effect on K_i. However, if firms have plants in multiple locations and cross-location spillovers are non-zero, as is true in our data, the optimal subsidy will not be uniform but plant-specific.

Second, Proposition 1 implies that the optimal plant-specific subsidy can be expressed using only a few elasticities ($\sigma, \alpha, \delta, \omega$), plant-level networks ($\mathbf{1}^J_{ni}$), and plant-level revenues as sufficient statistics. In particular, model primitives such as the costs of innovation investments or bilateral trade costs, or endogenous equilibrium objects such as prices, marginal costs, or location-wide knowledge, do not show up in equation (25).

6.3 Social versus Private Value of Innovation

A much studied object in the literature is the wedge between the social and private value of innovation. Our theoretical framework allows us to characterize this wedge at the location level accounting for all sources of inefficiencies in our model. We begin by defining the private and social value of innovation in a given location.

Definition 3. The *private value of innovation* in location i, β_i, is defined as the sum of plant-level private values, $\beta_i \equiv \sum_{J\ni i} \beta_{iJ} N_J$. The *social value of innovation* in location i, γ_i, is defined as the sum of plant-level social values, $\gamma_i \equiv \sum_{J \ni i} \gamma_{iJ} N_J$.

At the plant level, γ_{iJ}/β_{iJ} captures the wedge between social and private returns to innovation. At the location level, γ_i/β_i captures the average wedge between social and private returns across all plants in a location, weighted by plant-level innovation expenditures:

$$\frac{\gamma_i}{\beta_i} = \sum_{J \ni i} \frac{\gamma_{iJ}}{\beta_{iJ}} \frac{N_J h_{iJ}}{\sum_{J' \ni i} N_{J'} h_{iJ'}}.$$

We refer to γ_i/β_i as "*social-private innovation wedge*." A location has a higher wedge if private incentives lead to larger under-investment in innovation relative to the social value.

We use boldface notation to denote vectors and matrices. Let $\boldsymbol{I}$ denote the identity matrix and $\Omega \equiv [\omega_{ni}]$ denote the matrix with entries ω_{ni}. Let $s_{ni} \equiv \frac{\sum_J N_J r_{nJ} \mathbf{1}_{ni}^J}{\sum_J N_J r_{nJ}}$ denote the revenue-weighted fraction of plants in location n that are connected to location i through plant-level networks, and let $\boldsymbol{S} \equiv [s_{ni}]$ denote the corresponding matrix form. Let $\boldsymbol{r} \equiv \left[\sum_J N_J r_{iJ}\right]$ denote the column vector of total plant-level revenues in each location, and let $\boldsymbol{\gamma} \equiv [\gamma_i]$ and $\boldsymbol{\beta} \equiv [\beta_i]$. Our next result characterizes $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ using sufficient statistics that can be readily measured in the data.

Proposition 2. *(i)* $\boldsymbol{\beta}' = \frac{\alpha}{\sigma} \boldsymbol{r}' (\Omega \circ \boldsymbol{S})$; *(ii)* $\boldsymbol{\gamma}' = \frac{\sigma}{\sigma - 1} \boldsymbol{\beta}' (\boldsymbol{I} - \delta \Omega \circ \boldsymbol{S})^{-1}$, *where* $\circ$ *denotes the Hadamard product.*

Proof. See Online Appendix B.3. □

The first part of Proposition 2 characterizes the private value of innovation in a given location. Consider again first the special case when either cross-location spillovers are zero (so that Ω is the identity matrix $\boldsymbol{I}$) or all firms are single-location firms (so that $\boldsymbol{S}$ is the identity matrix). In either case, the (Hadamard) product $\Omega \circ \boldsymbol{S}$ is equal to the identity matrix. Proposition 2 then implies that the private value of innovation in a given location is $\beta_i = \frac{\alpha}{\sigma} r_i$. By contrast, when firms have plants in multiple locations ($\boldsymbol{S} \neq \boldsymbol{I}$) and cross-location spillovers are non-zero ($\Omega \neq \boldsymbol{I}$), the first part of Proposition 2 implies that $\beta_n = \alpha \sum_i \pi_i \omega_{ni} s_{ni}$. Thus, firms have stronger private incentives to invest in innovation in a given location if the investments generate larger spillovers (ω_{ni}) on other locations to which they have strong connections (s_{ni}) and which are highly profitable (π_i).

As for the second part of Proposition 2, we can rewrite $\boldsymbol{\gamma}$ as:

$$\begin{aligned}\boldsymbol{\gamma}' &= \frac{\sigma}{\sigma-1}\boldsymbol{\beta}'\left(\boldsymbol{I}-\delta\Omega\circ\boldsymbol{S}\right)^{-1} \\ &= \frac{\sigma}{\sigma-1}\boldsymbol{\beta}'\left(\boldsymbol{I}+\delta\Omega\circ\boldsymbol{S}+\left(\delta\Omega\circ\boldsymbol{S}\right)^{2}+\left(\delta\Omega\circ\boldsymbol{S}\right)^{3}+\cdots\right). \qquad (27)\end{aligned}$$

The scalar $\frac{\sigma}{\sigma-1}$ captures that—as in equation (19)—the impact of innovation investments on social welfare is $\frac{\sigma}{\sigma-1}$ times firm profits. The Leontief-inverse $\left(\boldsymbol{I}-\delta\Omega\circ\boldsymbol{S}\right)^{-1}$ captures the network effects of knowledge spillovers that propagate across locations through the plant-level networks of multi-location firms. To understand this network propagation effect, consider again first the special case when either cross-location spillovers are zero or all firms are single-location firms. Equation (27) then implies that the social-private innovation wedge is constant across all locations and equal to:

$$\frac{\gamma_i}{\beta_i}=\frac{\sigma}{\sigma-1}\frac{1}{1-\delta}, \qquad (28)$$

which coincides with the optimal innovation subsidy in (26).

In the general case where firms have plants in multiple locations and cross-location spillovers are non-zero, firms additionally under-invest in innovation because they now also ignore the impacts of their innovation investments on other firms' plants in other locations, via cross-location knowledge spillovers. These spatial spillover effects are captured by the cross-location elasticity matrix $\boldsymbol{\Omega}$ and are more socially valuable when a greater (revenue-weighted) fraction of plants in a given location are connected to other locations with stronger private incentives, as reflected by those locations' $\boldsymbol{\beta_i}$ values.[24]

Finally, note that both the vector $\boldsymbol{r}$, which measures the total revenue among all plants in a given location, and the matrix $\boldsymbol{S}$, which provides the revenue-weighted fraction of plants in a given location that are connected to a given other location, can be easily measured in Census data. Hence, by calibrating only two parameters, δ and ω, we can recover the relative rankings of locations based on their social-private innovation wedges (the scalars σ and α are not needed for such relative rankings).

[24] In Online Appendix C.2, we derive the analogue of Proposition 2 in an extended version of our model with heterogeneous productivities and endogenous innovation decisions along the extensive margin, so that ultimately only a subset of the plants innovates in equilibrium.

6.4 A Three-Location Example

Consider a setting with three locations $\{a, b, c\}$ and six firms. Each location has three plants. The firm ownership of the plants and their distribution across locations are as follows:

$$\begin{array}{cc} & \text{Firm} \\ & \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \text{Location} \begin{array}{c} a \\ b \\ c \end{array} & \begin{bmatrix} \checkmark & \checkmark & \checkmark & & & \\ \checkmark & \checkmark & & \checkmark & & \\ & & \checkmark & & \checkmark & \checkmark \end{bmatrix} \end{array}$$

For simplicity, assume all plants have the same revenue. Given the distribution and firm ownership of the plants, the matrix $\boldsymbol{S}$, which shows the revenue-weighted fraction of plants in a given location that are connected to a given other location, is:

$$\boldsymbol{S} = \begin{bmatrix} 1 & 2/3 & 1/3 \\ 2/3 & 1 & 0 \\ 1/3 & 0 & 1 \end{bmatrix}.$$

By construction, all three locations have the same number of plants and the same revenue. The locations differ only in their connectedness through the firms' networks of plants. Specifically, locations a and b are connected through the plants of firms 1 and 2, locations a and c are connected through the plants of firm 3, and locations b and c are not directly connected.

We can rank the three locations based on their social-private innovation wedges as follows:

$$\frac{\gamma_a}{\beta_a} > \frac{\gamma_b}{\beta_b} > \frac{\gamma_c}{\beta_c}.$$

The social-private innovation wedge is largest in location a and smallest in location c. Location a has the highest wedge because innovation there creates spillovers to both locations b and c. Location b has a higher wedge than location c because the spillovers from location b to location a occur through two firms, 1 and 2, whereas the spillovers from location c to location a only occur through a single firm, 3.

7 The Social-Private Innovation Wedge

7.1 Taking the Model to the Data

Our model provides a characterization of the wedge between the social and private returns to innovation at the location level ("social-private innovation wedge"). This wedge can be described in terms of a simple sufficient statistic, γ_n/β_n. In Section 7.2 we rank all U.S. tech clusters based on this wedge, and in Section 7.3 we examine how an increase in the interconnectedness of tech clusters affects the social-private innovation wedge in each location. In this section, we describe how we can compute this wedge using merged USPTO–U.S. Census Bureau micro data.

There are 179 locations. Accordingly, in Proposition 2, $\boldsymbol{\gamma}' = \frac{\sigma}{\sigma-1}\boldsymbol{\beta}'\left(\boldsymbol{I} - \delta\Omega \circ \boldsymbol{S}\right)^{-1}$ and $\boldsymbol{\beta}' = \frac{\alpha}{\sigma}\boldsymbol{r}'\left(\Omega \circ \boldsymbol{S}\right)$ are 179×1 vectors with entries γ_n and β_n, respectively. We compute $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ for each year and take averages across all years. (The scalars σ and α are the same for all locations and drop out when computing the ratio between any two wedges.) In equation (10), δ captures the strength of local knowledge spillovers (recall that $\omega_{ni} = 1$ if $n = i$). For this, we use the estimated within-cluster elasticity fom column (3) of Table 2. Similarly, $\delta\omega$ captures the strength of cross-location knowledge spillovers. For this, we use the estimated cross-cluster elasticity fom column (5) of Table 3. Both elasticities pertain to plant-level inventor productivity; we obtain similar results if we use the corresponding TFP elasticities. To obtain ω (to populate the matrix Ω), we divide $\delta\omega$ by δ. In reduced form, this means we divide the cross-cluster elasticity by the within-cluster elasticity. The matrix Ω is a 179×179 matrix with diagonals of 1 and off-diagonals of ω (see Section 6.1); the latter are scaled by the average number of connected clusters, which is 3.22 (see Table 1). If a location n has no connected clusters in location i, we set the corresponding $\{n, i\}$ entry in the Ω matrix to zero.

The matrix $\boldsymbol{S}$ is a 179×179 matrix with entries $s_{ni} \equiv \frac{\sum_J N_J r_{nJ} \mathbf{1}_{ni}^J}{\sum_J N_J r_{nJ}}$ representing the revenue-weighted fraction of innovating plants in location n that have a connected cluster in location i. To compute this matrix, we first construct an auxiliary location $\times$ firm matrix similar to the one in the three-location example, where each entry r_{nJ} represents the total value of plant-level revenues of firm J in location n. We obtain plant-level revenues directly from the Census data ("shipments" in the ASM and CMF). Summing over all entries in a given row (i.e., location) n yields the total value of plant-level revenues in location n, $\sum_J N_J r_{nJ}$. Similarly, summing over only those entries in row n that are associated with firms J that also have entries in row i yields $\sum_J N_J r_{nJ} \mathbf{1}_{ni}^J$. Dividing $\sum_J N_J r_{nJ} \mathbf{1}_{ni}^J$ by $\sum_J N_J r_{nJ}$ yields the entry s_{ni} in the $\boldsymbol{S}$ matrix. Also, $\sum_J N_J r_{nJ}$ is the entry in row n of the vector $\boldsymbol{r}$, which is a 179×1 vector representing the total value of plant-level revenues in each location.

7.2 Ranking U.S. Tech Clusters

We rank all 179 cities based on their values of γ_n/β_n. Since this is too large a ranking to report, Table 8 only shows the top and bottom ten clusters with the highest and lowest values of γ_n/β_n, respectively. For comparison, the table also shows the top and bottom ten clusters ranked by size (number of inventors). What stands out, first and foremost, is that larger clusters have a higher social-private innovation wedge. Three out of the top ten clusters ranked by this wedge are also in the top ten ranked by size (marked in red): Los Angeles-Long Beach-Riverside, San Jose-San Francisco-Oakland, and Chicago-Naperville-Michigan City. Moreover, seven out of the top ten clusters with the highest wedge are in the top decile ranked by size, and all ten are in the top quartile.

While there is a clear association between the social-private innovation wedge and cluster size, it is far from perfect. Indeed, seven out of the top ten clusters ranked by this wedge are *not* in the top ten ranked by size. Thus, cluster size per se is not what is driving the ranking. Granted, larger clusters have more local knowledge spillovers and thus a higher social return to innovation. But firms' private incentives to invest are also greater in larger clusters, in proportion to the social return. As a result, if it was only for local knowledge spillovers, the social-private innovation wedge would be *invariant* to cluster size (c.f. equation (28) showing that, without cross-location spillovers, the social-private innovation wedge is constant across locations). This invariance echoes a well-known result going back to Glaeser and Gottlieb (2009) stating that place-based redistributions are not welfare-enhancing unless the elasticity of productivity to agglomeration differs across locations. Hence, what matters for the optimality of policy intervention is the *elasticity* with respect to agglomeration, not the size of the local economy.

What is instead driving the location ranking—and breaking the Glaeser-Gottlieb logic in our case—is the extent to which clusters are connected to other (especially productive and well-connected) clusters. In the presence of cross-cluster innovation spillovers, firms' innovation investments raise the productivity of other firms' inventors and plants in connected clusters. Since firms do not internalize these spillovers, clusters that are well-connected to other clusters have a high social-private innovation wedge. In our data, the correlation between γ_n/β_n and a simple empirical proxy for a cluster's connectedness—the average number of connected clusters across all plants in a cluster—is 41.1%.[25] Accordingly, innovation policy should focus on well-connected clusters, from

[25] As well-connected clusters are typically also larger, there is a positive correlation between γ_n/β_n and a

where innovations spill over to other, well-connected clusters, and so on.

7.3 Increasing the Interconnectedness of U.S. Tech Clusters

As clusters are becoming more interconnected, the social returns to innovation increase. But so do firms' private incentives to innovate. The relative effect, captured by γ_n/β_n, is theoretically ambiguous. To explore the effect of an increase in the interconnectedness of clusters on the social-private innovation wedge, we again take our model to the data. Specifically, we counterfactually increase all off-diagonal elements of the $\boldsymbol{S}$ matrix by 10% and compute the percentage change in γ_n/β_n for all 179 cities. (Recall that an element of the $\boldsymbol{S}$ matrix represents the revenue-weighted fraction of innovating plants in location n that have a connected cluster in location i.)

While the effect of an increase in interconnectedness is theoretically ambiguous, it is clear in the data: a 10% increase in the off-diagonals of the $\boldsymbol{S}$ matrix raises γ_n/β_n by 2.6% on average. Hence, as clusters are becoming more connected with each other, the gap between the social and private returns to innovation widens, exacerbating the under-investment problem. However, while γ_n/β_n increases by 2.6% on average, there is significant heterogeneity across locations. At the 25th percentile, γ_n/β_n only increases by 0.8%, and a few clusters even experience a slight *decrease* in γ_n/β_n. On the other hand, clusters like Chicago-Naperville-Michigan City or Washington-Baltimore-Northern Virginia experience substantial increases of 7.2% and 6.6%, respectively. The median increase is 1.9%, which is well below the average.

Table 9 shows the top and bottom ten clusters with the highest and lowest percent change in γ_n/β_n, respectively. For comparison, the table also shows the top and bottom ten clusters ranked by size (number of inventors). Similar to Table 8, there is a clear visual association between the change in the social-private innovation wedge and cluster size. Two out of the top ten clusters with the highest increase in the wedge are also in the top ten ranked by size (marked in red): Chicago-Naperville-Michigan City and Boston-Worcester-Manchester. Across all 179 cities, the correlation between the change in the wedge and cluster size is 28.3%, and the correlation with the average number of connected clusters—a simple empirical proxy for a cluster's connectedness—is 22.4%. Accordingly, an increase in the interconnectedness of clusters especially raises the social-private innovation wedge in large and well-connected clusters.

---

cluster's own size. However, this correlation is only 16.6% and thus much weaker than the correlation with a cluster's connectedness.

8 Conclusion

Within the firm, knowledge diffusion need not rely on face-to-face interactions but can take place through conventional channels, such as emails, internal memos, or video conferencing. If knowledge spreads locally across inventors of different firms, we would therefore expect these inventors to pass on this knowledge to other inventors and plants of the same firm, across different locations. In other words, we would expect the same economic forces that generate innovation spillovers *within* a tech cluster also to generate innovation spillovers *across* tech clusters.

We test this hypothesis using merged USPTO–U.S. Census Bureau micro data. Cross-cluster innovation spillovers are identified through within-plant variation in the size of a plant's "connected clusters"—the number of other firms' inventors in the plant's research field in other cities where the plant's parent firm also has innovating plants—in combination with granular city $\times$ field $\times$ year fixed effects. Moreover, we employ an IV design that isolates variation in the size of a plant's connected clusters that plausibly originates from outside those clusters. We find robust evidence of cross-cluster innovation spillovers. Consistent with a knowledge diffusion channel, these spillovers do not decay with distance, and plants cite disproportionately more patents by other firms in connected clusters as the size of those clusters increases.

Firms are unlikely to fully internalize the impacts of their innovative activities on other firms' innovation and productivity. To inform policy, we develop a tractable model of spatial innovation that features both within- and cross-cluster innovation spillovers. We show that the optimal government policy is plant-specific and disproportionately targets plants that are well-connected to other clusters, especially clusters that are in turn well-connected to other clusters, thus generating the biggest "bang for the buck." Importantly, we derive a sufficient statistic—the "social-private innovation wedge"—that captures the wedge between the social and private returns to innovation in a location. Taking our model to the data, we rank all U.S. tech clusters based on this statistic. We find that larger clusters exhibit a greater wedge between social and private returns; however, this is not because of local knowledge spillovers, but because larger clusters are more connected to other clusters through firms' networks of innovating plants.

References

Aghion, Philippe, John Van Reenen, and Luigi Zingales, 2013, Innovation and Institutional Ownership, American Economic Review 103, 277–304.

Amore, Mario Daniele, Cédric Schneider, and Alminas Žaldokas, 2013, Credit Supply and Corporate Innovation, Journal of Financial Economics 109, 835–855.

Arqué-Castells, Pere, and Daniel Spulber, 2022, Measuring the Private and Social Returns to R&D: Unintended Spillovers versus Technology Markets, Journal of Political Economy 130, 1860–1918.

Atkin, David, Keith Chen, and Anton Popov, 2022, The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley, NBER Working Paper 30147.

Audretsch, David, and Maryann Feldman, 2004, Knowledge Spillovers and the Geography of Innovation, in: Handbook of Regional and Urban Economics, Volume 4, Vernon Henderson and Jacques Thisse (eds.), 2713–2739. Amsterdam: Elsevier.

Bena, Jan, Serdar Dinc, and Isil Erel, 2022, The International Propagation of Economic Downturns through Multinational Companies: The Real Economy Channel, Journal of Financial Economics 146, 277–304.

Bernstein, Shai, 2015, Does Going Public Affect Innovation? Journal of Finance 70, 1365–1403.

Biermann, Marcus, and Kilian Huber, 2024, Tracing the International Transmission of a Crisis through Multinational Firms, Journal of Finance 79, 1789–1829.

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

Bloom, Nicholas, John Van Reenen, and Heidi Williams, 2019, A Toolkit of Policies to Promote Innovation, Journal of Economic Perspectives 33, 163–184.

Carlino, Gerald, and William Kerr, 2015, Agglomeration and Innovation, in: Handbook of Regional and Urban Economics, Volume 5, Gilles Duranton, Vernon Henderson, and William Strange (eds.), 349–404. Amsterdam: Elsevier.

Cohen, Wesley, and Daniel Levinthal, 1989, The Two Faces of R&D, Economic Journal 99, 569–596.

Cornaggia, Jess, Yifei Mao, Xuan Tian, and Brian Wolfe, 2015, Does Banking Competition Affect Innovation? Journal of Financial Economics 115, 189–209.

Cravino, Javier, and Andrei Levchenko, 2017, Multinational Firms and International Business Cycle Transmission, Quarterly Journal of Economics 132, 921–962.

Derrien, François, Ambrus Kecskés, and Phuong-Anh Nguyen, 2023, Labor Force Demographics and Corporate Innovation, Review of Financial Studies 36, 2797–2838.

Giroud, Xavier, and Holger Mueller, 2019, Firm's Internal Networks and Local Economic Shocks, American Economic Review 109, 3617–3649.

Giroud, Xavier, Simone Lenzu, Quinn Maingi, and Holger Mueller, 2024, Propagation and Amplification of Local Productivity Spillovers, Econometrica 92, 1589–1619.

Glaeser, Edward, and Joshua Gottlieb, 2009, The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States, Journal of Economic Literature 47, 983–1028.

Griliches, Zvi, 1979, Issues in Assessing the Contribution of Research and Development to Productivity Growth, Bell Journal of Economics 10, 92–116.

Gruber, Jonathan, Enrico Moretti, and Daniel Wilson, 2023, Place-Based Productivity and Costs in Science, in: Entrepreneurship and Innovation Policy and the Economy, Volume 2, Benjamin Jones and Josh Lerner (eds.), 167–184, Chicago: University of Chicago Press.

Hombert, Johan, and Adrien Matray, 2017, The Real Effects of Lending Relationships on Innovative Firms and Inventor Mobility, Review of Financial Studies 30, 2413–2445.

Jaffee, Adam, 1986, Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value, American Economic Review 76, 984–1001.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, 1993, Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, Quarterly Journal of Economics 108, 577–598.

Kerr, William, and Frederic Robert-Nicoud, 2020, Tech Clusters, Journal of Economic Perspectives 34, 50–76.

Li, Xuelin, Tong Liu, and Lucian Taylor, 2023, Common Ownership and Innovation Efficiency, Journal of Financial Economics 147, 475–497.

Liu, Ernest, and Song Ma, 2024, Innovation Networks and R&D Allocation, mimeo, Princeton University.

Markusen, James, 1984, Multinationals, Multi-Plant Economies, and the Gains from Trade, Journal of International Economics 16, 205–226.

Matray, Adrien, 2021, The Local Innovation Spillovers of Listed Firms, Journal of Financial Economics 141, 395–412.

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

Moretti, Enrico, and Daniel Wilson, 2017, The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists, American Economic Review 107, 1858–1903.

Saxenian, AnnaLee, 1994, Regional Advantage: Culture and Competition in Silicon Valley and Route 128, Harvard University Press.

Seru, Amit, 2014, Firm Boundaries Matter: Evidence from Conglomerates and R&D Activity, Journal of Financial Economics 111, 381–405.

Table 1: Descriptive Statistics

This table presents plant- and firm-level statistics for the main analysis sample consisting of 57,000 plant-year observations and 11,000 firm-year observations. Patents is the total number of patents (in the plant's field) per plant and year. Connected clusters is the number of other cities in which the firm has plants (in the plant's field) with inventors. Connected cluster size is the total number of other firms' inventors (in the plant's field) across all connected clusters. Cities are BEA economic areas. Fields are based on 1-digit CPC codes. Inventor share at the firm-plant level is the ratio of the firm's plants with inventors to all plants of the firm, averaged across all firms. Inventor share at the firm-county (city, state) level is the ratio of counties (cities, states) in which the firm has plants with inventors to all counties (cities, states) in which the firm has plants, averaged across all firms. The sample period is from 1976 to 2018.

	(1)	(2)
	Mean	Std. dev.
Plants:		
Employees	699	1,623
Shipments ($M)	275	1,670
Inventors	4.99	13.67
Patents	10.87	43.07
Connected Clusters	3.22	2.81
Connected Cluster Size	1,915	2,748
Firms:		
Employees	6,853	13,280
Shipments ($M)	2,524	6,730
Plants	17.51	23.19
Counties	15.23	18.47
Cities	11.77	11.72
States	9.16	7.33
Inventor Share:		
Plants	0.56	0.31
Counties	0.55	0.31
Cities	0.52	0.33
States	0.49	0.34

Table 2: Within-Cluster Innovation Spillovers

This table presents variants of the regression in equation (1) in which the dependent variable is either the ratio (in logs) of the number of patents to the number of inventors (all in the plant's field) at the plant level or plant-level TFP. Cluster size is the number of other firms' inventors in the plant's city and field. In column (5), the sample consists of plants without inventors. For these plants, fixed effects are based on the firm's main field, and cluster size is the total number of other firms' inventors in the plant's city in the firm's main field. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP	TFP
	(1)	(2)	(3)	(4)	(5)
Log(Cluster Size)	0.078***	0.029***	0.076***	0.023***	0.003
	(0.010)	(0.004)	(0.013)	(0.007)	(0.003)
Field FE	Yes	Yes	-	-	-
Year FE	Yes	Yes	-	-	-
Field × Year FE	No	No	Yes	Yes	Yes
City × Year FE	No	No	Yes	Yes	Yes
City × Field FE	No	No	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes
Observations	134,000	134,000	134,000	134,000	259,000
R-squared	0.39	0.67	0.44	0.70	0.74

Table 3: Cross-Cluster Innovation Spillovers

This table presents variants of the regression in equation (2) in which the dependent variable is either the ratio (in logs) of the number of patents to the number of inventors (all in the plant's field) at the plant level or plant-level TFP. (Connected) cluster size is the total number of other firms' inventors (in the plant's field) across all other cities in which the firm has plants (in the plant's field) with inventors. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Cluster Size)	0.024***	0.010***	0.023***	0.012***	0.021***	0.012***
	(0.005)	(0.003)	(0.007)	(0.004)	(0.007)	(0.004)
Field FE	Yes	Yes	-	-	-	-
Year FE	Yes	Yes	-	-	-	-
Field × Year FE	No	No	Yes	Yes	-	-
City × Year FE	No	No	Yes	Yes	-	-
City × Field FE	No	No	Yes	Yes	-	-
City × Field × Year FE	No	No	No	No	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	57,000	57,000	57,000	57,000	57,000	57,000
R-squared	0.45	0.69	0.52	0.75	0.59	0.79

Table 4: Non-Innovating Plants

This table presents variants of the regressions in columns (5) and (6) of Table 3. In columns (1) and (2), (connected) cluster size is the total number of other firms' inventors (in the plant's field) across all other cities in which the firm has plants but no inventors. In column (3), the sample consists of plants without inventors. For these plants, fixed effects are based on the firm's main field, and (connected) cluster size is the total number of other firms' inventors (in the firm's main field) across all other cities in which the firm has plants (in the firm's main field) with inventors. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Pat/Inv)	TFP	TFP
	(1)	(2)	(3)
Log(Cluster Size)	0.002	0.001	0.006**
	(0.008)	(0.005)	(0.003)
City × Field × Year FE	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes
Observations	80,000	80,000	227,000
R-squared	0.52	0.75	0.82

Table 5: Instrumental Variable Estimation

This table presents variants of the regressions in columns (5) and (6) of Table 3 estimated by 2SLS. The excluded instrument, Z, is the number of inventors (in the plant's field) in other cities where the plant's parent firm has no presence and which are employed by other firms that have inventors in the plant's connected clusters but not in the plant's city. Column (1) shows the first stage; columns (2) and (3) shows 2SLS estimates. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Cluster Size)	Log(Pat/Inv)	TFP
	(1)	(2)	(3)
Z	0.107***		
	(0.013)		
Log(Cluster Size)		0.023**	0.014**
		(0.011)	(0.007)
City × Field × Year FE	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes
Observations	57,000	57,000	57,000
R-squared	0.85	0.59	0.79
F-statistic	70.60		

Table 6: Geographical Distance

This table presents variants of the regressions in columns (5) and (6) of Table 3 in which connected clusters within a 100, 250, or 500 mile radius around the plant are excluded. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Pat/Inv)			TFP		
	(1)	(2)	(3)	(4)	(5)	(6)
	Excluding clusters within 100 miles	Excluding clusters within 250 miles	Excluding clusters within 500 miles	Excluding clusters within 100 miles	Excluding clusters within 250 miles	Excluding clusters within 500 miles
Log(Cluster Size)	0.021***	0.020***	0.019**	0.011***	0.011**	0.011**
	(0.007)	(0.007)	(0.008)	(0.004)	(0.005)	(0.005)
City × Field × Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53,000	46,000	35,000	53,000	46,000	35,000
R-squared	0.60	0.60	0.61	0.79	0.80	0.82

Table 7: Patent Citations

This table presents variants of the regressions in Table 3 and columns (1) and (2) of Table 4 in which the dependent variable is the share of the plant's citations to other firms' patents from connected clusters relative to all its citations to other firms' patents (in the plant's field). In column (4), citations to patents filed in year *t* are excluded. In column (5), (connected) cluster size is the total number of other firms' inventors (in the plant's field) across all other cities in which the firm has plants but no inventors. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Patent Citation Share				
	(1)	(2)	(3)	(4)	(5)
Log(Cluster Size)	0.018**	0.022***	0.019**	0.019**	0.002
	(0.007)	(0.008)	(0.008)	(0.008)	(0.010)
Field FE	Yes	-	-	-	-
Year FE	Yes	-	-	-	-
Field × Year FE	No	Yes	-	-	-
City × Year FE	No	Yes	-	-	-
City × Field FE	No	Yes	-	-	-
City × Field × Year FE	No	No	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes
Observations	57,000	57,000	57,000	57,000	80,000
R-squared	0.55	0.63	0.69	0.69	0.58

Table 8: Ranking U.S. Tech Clusters

This table ranks U.S. tech clusters based on their social-private innovation wedge, γ_n/β_n (left) and cluster size (right). Clusters are the 179 BEA economic areas ("cities"). Cluster size is the number of inventors in a city. Clusters are listed in ascending order.

Top 10 (Social-Private Innovation Wedge)	Top 10 (Cluster Size)
Atlanta-Sandy Springs-Gainesville, GA-AL	San Jose-San Francisco-Oakland, CA
Denver-Aurora-Boulder, CO	New York-Newark-Bridgeport, NY-NJ-CT-PA
Washington-Baltimore-Northern Virginia, DC	Minneapolis-St. Paul-St. Cloud, MN-WI
Los Angeles-Long Beach-Riverside, CA	Detroit-Warren-Flint, MI
San Jose-San Francisco-Oakland, CA	Boston-Worcester-Manchester, MA-NH
Philadelphia-Camden-Vineland, PA-NJ-DE-MD	Los Angeles-Long Beach-Riverside, CA
Chicago-Naperville-Michigan City, IL-IN-WI	Chicago-Naperville-Michigan City, IL-IN-WI
Phoenix-Mesa-Scottsdale, AZ	Portland-Vancouver-Beaverton, OR-WA
Raleigh-Durham-Cary, NC	Dallas-Fort Worth, TX
Houston-Baytown-Huntsville, TX	Cleveland-Akron-Elyria, OH
Bottom 10 (Social-Private Innovation Wedge)	**Bottom 10 (Cluster Size)**
Dothan-Enterprise-Ozark, AL	San Angelo, TX
Cape Girardeau-Jackson, MO-IL	Rapid City, SD
Farmington, NM	Pendleton-Hermiston, OR
Pendleton-Hermiston, OR	Honolulu, HI
Alpena, MI	Abilene, TX
Bismarck, ND	Aberdeen, SD
Minot, ND	Minot, ND
Bangor, ME	Great Falls, MT
Scotts Bluff, NE	Anchorage, AK
Great Falls, MT	Santa Fe-Espanola, NM

Table 9: Increasing the Interconnectedness of U.S. Tech Clusters

This table ranks U.S. tech clusters based on the percent change in their social-private innovation wedge following a 10% increase in the off-diagonal elements of the **S** matrix (left) and cluster size (right). Clusters are the 179 BEA economic areas ("cities"). Cluster size is the number of inventors in a city. Clusters are listed in ascending order.

Top 10 (Change in Wedge)	Top 10 (Cluster Size)
Chicago-Naperville-Michigan City, IL-IN-WI (7.2%)	San Jose-San Francisco-Oakland, CA
Washington-Baltimore-Northern Virginia, DC (6.6%)	New York-Newark-Bridgeport, NY-NJ-CT-PA
Austin-Round Rock, TX (6.2%)	Minneapolis-St. Paul-St. Cloud, MN-WI
Boston-Worcester-Manchester, MA-NH (6.1%)	Detroit-Warren-Flint, MI
Indianapolis-Anderson-Columbus, IN (6.1%)	Boston-Worcester-Manchester, MA-NH
Houston-Baytown-Huntsville, TX (5.8%)	Los Angeles-Long Beach-Riverside, CA
San Diego-Carlsbad-San Marcos, CA (5.6%)	Chicago-Naperville-Michigan City, IL-IN-WI
Rochester-Batavia-Seneca Falls, NY (5.5%)	Portland-Vancouver-Beaverton, OR-WA
Seattle-Tacoma-Olympia, WA (5.3%)	Dallas-Fort Worth, TX
Pittsburgh-New Castle, PA (5.3%)	Cleveland-Akron-Elyria, OH

Bottom 10 (Change in Wedge)	Bottom 10 (Cluster Size)
Amarillo, TX (0.4%)	San Angelo, TX
Flagstaff, AZ (0.4%)	Rapid City, SD
Abilene, TX (0.4%)	Pendleton-Hermiston, OR
Kearney, NE (0.3%)	Honolulu, HI
Alpena, MI (0.2%)	Abilene, TX
Gulfport-Biloxi-Pascagoula, MS (0.0%)	Aberdeen, SD
Jonesboro, AR (-0.0%)	Minot, ND
Fort Smith, AR-OK (-0.1%)	Great Falls, MT
Paducah, KY-IL (-0.2%)	Anchorage, AK
Pueblo, CO (-0.2%)	Santa Fe-Espanola, NM

Innovation Spillovers across U.S. Tech Clusters

Xavier Giroud   Ernest Liu   Holger Mueller

Online Appendix

A Additional Figures and Tables

Figure A.1: BEA Economic Areas

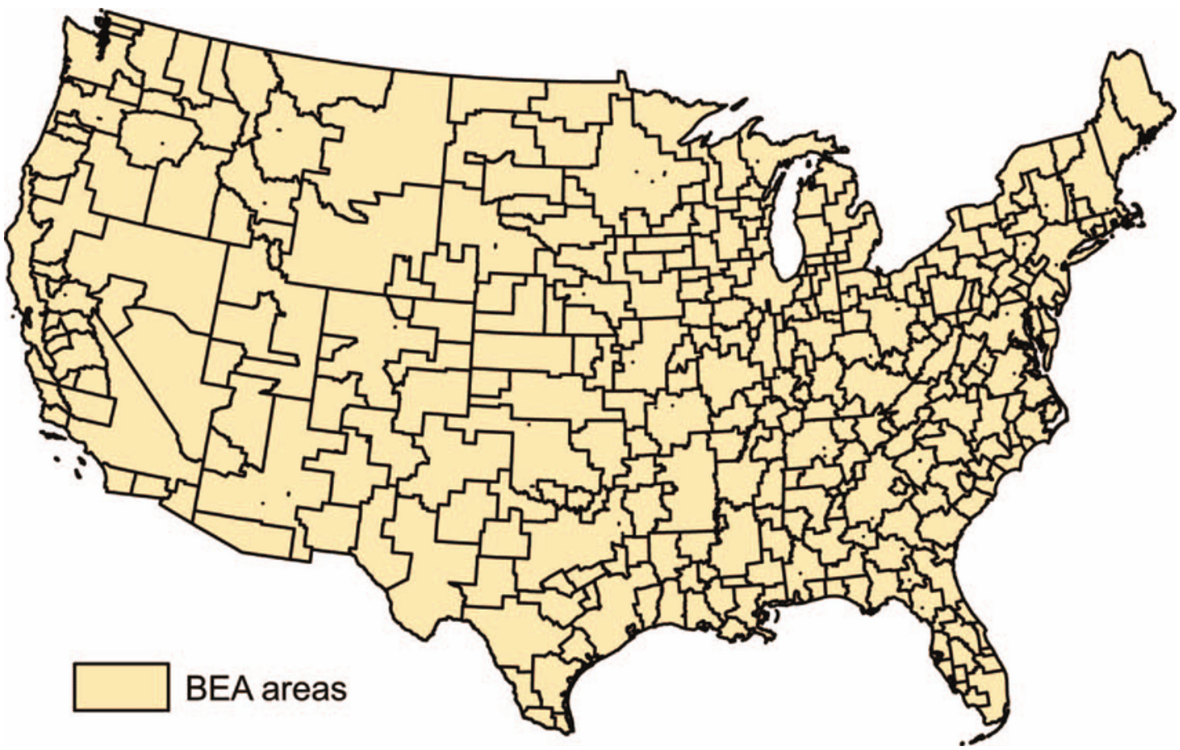

Table A.1: Robustness

This table presents variants of the regressions in columns (5) and (6) of Table 3. In columns (1) and (2), inventors are assigned to the nearest firm plant in the inventor's city. In columns (3) and (4), a broader concept of cluster size is employed, where it is assumed that inventors work in a cluster also in between patent filings. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP
	(1)	(2)	(3)	(4)
Log(Cluster Size)	0.022***	0.011***	0.018***	0.010***
	(0.005)	(0.004)	(0.006)	(0.003)
City × Field × Year FE	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes
Observations	70,000	70,000	57,000	57,000
R-squared	0.56	0.77	0.59	0.79

Table A.2: Robustness (Continued)

This table presents variants of the regressions in columns (5) and (6) of Table 3. In columns (1) and (2), the plant's "field" is based on all fields in which it has patents, and its patent output, inventors, and connected clusters are all based on those fields. In columns (3) and (4), the city × field × year fixed effects are replaced with city × class × year fixed effects, where technology classes are defined at the 3-digit CPC code level. In column (5), the sample includes all LBD establishments with inventor matches. Observations are weighted by plant (establishment) employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP	Log(Pat/Inv)
	(1)	(2)	(3)	(4)	(5)
Log(Cluster Size)	0.018***	0.010**	0.018**	0.010**	0.018***
	(0.006)	(0.004)	(0.008)	(0.005)	(0.003)
City × Field × Year FE	Yes	Yes	-	-	Yes
City × Class × Year FE	No	No	Yes	Yes	No
Plant FE	Yes	Yes	Yes	Yes	Yes
Observations	57,000	57,000	57,000	57,000	383,000
R-squared	0.61	0.80	0.66	0.84	0.71

Table A.3: Number of Plant Inventors

This table presents variants of the regressions in columns (3) and (4) of Table 2 and columns (5) and (6) of Table 3 in which (connected) cluster size is interacted with the number of plant inventors (in logs). In columns (1) to (4), number of inventors refers to the focal plant. In columns (5) and (6), number of inventors refers to the firm's innovating plants in connected clusters. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Within-Cluster		Cross-Cluster			
	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Cluster Size)	0.054***	0.015**	0.017**	0.009**	0.018**	0.009**
	(0.013)	(0.006)	(0.007)	(0.004)	(0.008)	(0.004)
Log(Cluster Size) × Log(Inventors)	0.025***	0.012***	0.004**	0.003**	0.002**	0.002**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Log(Inventors)	0.101**	0.010***	0.005**	0.008**	0.003**	0.004*
	(0.049)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Field × Year FE	Yes	Yes	-	-		
City × Year FE	Yes	Yes	-	-		
City × Field FE	Yes	Yes	-	-		
City × Field × Year FE	-	-	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	134,000	134,000	57,000	57,000	57,000	57,000
R-squared	0.44	0.70	0.59	0.79	0.59	0.79

Table A.4: Aggregation to the Cluster Level

This table presents variants of the regressions in columns (3) and (4) of Table 3 in which observations are aggregated at the city × field × year level. The dependent variable is either the ratio (in logs) of the total number of patents to the total number of inventors based on all plants in the cluster or the employment-weighted average TFP across all plants in the cluster. (Connected) cluster size is either the sum (columns (1) and (2)) or average (columns (3) and (4)) of the plant-level (connected) cluster sizes across all plants in the cluster. Observations are weighted by city × field employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Pat/Inv)	TFP	Log(Pat/Inv)	TFP
	(1)	(2)	(3)	(4)
Log(Cluster Size)	0.016**	0.010**	0.017**	0.011**
	(0.008)	(0.004)	(0.008)	(0.005)
Field × Year FE	Yes	Yes	Yes	Yes
City × Year FE	Yes	Yes	Yes	Yes
City × Field FE	Yes	Yes	Yes	Yes
Observations	18,000	18,000	18,000	18,000
R-squared	0.50	0.48	0.50	0.48

Table A.5: Patent Output

This table presents variants of the regressions in Table 3 and columns (1) and (2) of Table 4 in which the dependent variable is the log number of patents (in the plant's field) at the plant level. In column (4), (connected) cluster size is the total number of other firms' inventors (in the plant's field) across all other cities in which the firm has plants but no inventors. Observations are weighted by plant employment. Standard errors are double clustered at the city and year level. The sample period is from 1976 to 2018. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

	Log(Patents)			
	(1)	(2)	(3)	(4)
Log(Cluster Size)	0.036***	0.031***	0.030***	0.004
	(0.006)	(0.006)	(0.007)	(0.009)
Field FE	Yes	-	-	-
Year FE	Yes	-	-	-
Field × Year FE	No	Yes	-	-
City × Year FE	No	Yes	-	-
City × Field FE	No	Yes	-	-
City × Field × Year FE	No	No	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes
Observations	57,000	57,000	57,000	80,000
R-squared	0.63	0.69	0.74	0.61

B Proofs

B.1 Proof of Lemma 1

Let $x \equiv \left\{h_{iJ}, K_i\right\}_{i,J}$ and $\hat{x} \equiv \left\{\hat{h}_{iJ}, \hat{K}_i\right\}_{i,J}$ denote two equilibria, and let $\widetilde{y} \equiv \ln \hat{y} - \ln y$ for $y \in \left\{h_{iJ}, K_i\right\}_{i,J}$. Let $\zeta \equiv ||\ln x - \ln \hat{x}||_\infty$. In an equilibrium, $\left\{h_{iJ}, K_i\right\}$ must satisfy:

$$c_q h_{qJ} = \alpha \sum_n b_n \prod_{i \in J} \left(h_{iJ}^\alpha K_i^\delta\right)^{\omega_{ni}} \omega_{nq},$$

$$b_n = \sum_k \frac{1}{\sigma} \frac{\tau_{kn}^{1-\sigma}}{\sum_m \sum_{J' \ni m} N_{J'} \prod_{i \in J'} \left(h_{iJ'}^\alpha K_i^\delta\right)^{\omega_{mi}} \tau_{km}^{1-\sigma}},$$

and

$$K_n \equiv \sum_{J \ni n} N_J \prod_{i \in J} \left(h_{iJ}^\alpha K_i^\delta\right)^{\omega_{ni}}.$$

Define f_{qJ} and g_n as the equilibrium correspondence with:

$$h_{qJ} \equiv f_{qJ}\left(x\right)$$

and

$$K_n \equiv g_n\left(x\right).$$

By the intermediate value theorem, there must exist $x' \equiv \exp\left(t \ln x + (1-t)\ln \hat{x}\right)$ for $t \in [0,1]$ such that:

$$\widetilde{h}_{qJ} = \sum_{iJ'} \frac{\partial \ln f_{qJ}\left(x'\right)}{\partial \ln h_{iJ'}} \widetilde{h}_{iJ'} + \sum_i \frac{\partial \ln f_{qJ}\left(x'\right)}{\partial \ln K_i} \widetilde{K}_i$$

and

$$\widetilde{K} = \sum_{iJ'} \frac{\partial \ln g_n\left(x'\right)}{\partial \ln h_{iJ'}} \widetilde{h}_{iJ'} + \sum_i \frac{\partial \ln g_n\left(x'\right)}{\partial \ln K_i} \widetilde{K}_i.$$

Hence:

$$\begin{aligned}
\left|\widetilde{h}_{qJ}\right| &\leq \sum_{iJ'} \left|\frac{\partial \ln f_{qJ}\left(x'\right)}{\partial \ln h_{iJ'}}\right| \left|\widetilde{h}_{iJ'}\right| + \sum_i \left|\frac{\partial \ln f_{qJ}\left(x'\right)}{\partial \ln K_i}\right| \left|\widetilde{K}_i\right| \\
&\leq \left(\sum_{iJ'} \left|\frac{\partial \ln f_{qJ}\left(x'\right)}{\partial \ln h_{iJ'}}\right| + \sum_i \left|\frac{\partial \ln f_{qJ}\left(x'\right)}{\partial \ln K_i}\right|\right) \zeta \\
&\leq \left(\alpha + \delta\right)\left(1 + \left(\mathcal{J} - 1\right)\omega\right)\zeta
\end{aligned}$$

and

$$
\begin{aligned}
\left|\widetilde{K}\right| &\leq \sum_{iJ'}\left|\frac{\partial \ln g_n\left(x'\right)}{\partial \ln h_{iJ'}}\right|\left|\widetilde{h}_{iJ'}\right| + \sum_i \left|\frac{\partial \ln g_n\left(x'\right)}{\partial \ln K_i}\right|\left|\widetilde{K}_i\right| \\
&\leq \left(\sum_{iJ'}\left|\frac{\partial \ln g_n\left(x'\right)}{\partial \ln h_{iJ'}}\right| + \sum_i \left|\frac{\partial \ln g_n\left(x'\right)}{\partial \ln K_i}\right|\right)\zeta \\
&\leq (\alpha+\delta)\left(1+(\mathcal{J}-1)\,\omega\right)\zeta.
\end{aligned}
$$

We thus have $\zeta \leq (\alpha+\delta)\left(1+(\mathcal{J}-1)\,\omega\right)\zeta$. A sufficient condition for $\zeta = 0$ is that $(\alpha+\delta)\left(1+(\mathcal{J}-1)\,\omega\right) < 1$, establishing uniqueness.

B.2 Proof of Lemma 2

Equations (5) and (7) imply:

$$
y_{ni}\left(\nu\right) = \frac{\sigma-1}{\sigma}\frac{\left(\tau_{ni}\kappa_i\left(\nu\right)\right)^{-\sigma}}{\sum_i \int_\nu \left(\tau_{ni}\kappa_i\left(\nu\right)\right)^{1-\sigma} d\nu}.
$$

Substituting into the definition of Y_n in (3) and using the fact that $\kappa_{nJ}^{1-\sigma} = k_{nJ}$, we obtain:

$$
\begin{aligned}
\ln Y_n &= \ln\left[\sum_i \int y_{ni}\left(\nu\right)^{\frac{\sigma-1}{\sigma}} d\nu\right]^{\frac{\sigma}{\sigma-1}} \\
&= \frac{\sigma}{\sigma-1}\ln\left[\sum_i \int_\nu \left(\tau_{ni}\kappa_i\left(\nu\right)\right)^{1-\sigma} d\nu\right] - \ln\left(\sum_i \int_\nu \left(\tau_{ni}\kappa_i\left(\nu\right)\right)^{1-\sigma} d\nu\right) + \ln\frac{\sigma-1}{\sigma} \\
&= \frac{1}{\sigma-1}\ln\left[\sum_i \int_\nu \left(\tau_{ni}\kappa_i\left(\nu\right)\right)^{1-\sigma} d\nu\right] + \ln\frac{\sigma-1}{\sigma} \\
&= \frac{1}{\sigma-1}\ln\left[\sum_i \sum_{J\ni i} \tau_{ni}^{1-\sigma} N_J k_{iJ}\right] + \ln\frac{\sigma-1}{\sigma}.
\end{aligned}
$$

Taking the derivative with respect to k_{iJ}, summing across n, and substituting using the definition of b_i in (13), we have:

$$
\begin{aligned}
\sum_n \frac{\partial \ln Y_n}{\partial k_{iJ}} &= \sum_n \frac{1}{\sigma-1}\frac{N_J \tau_{ni}^{1-\sigma}}{\sum_m \sum_{J\ni m} \tau_{nm}^{1-\sigma} N_J k_{mJ}} \\
&= \frac{\sigma}{\sigma-1} N_J b_i.
\end{aligned}
$$

Lemma 2 then follows immediately from the definitions of $\mathcal{W}$ and $\hat{\mathcal{W}}$.

B.3 Proof of Proposition 2

Note that:

$$\beta_i \equiv \textstyle\sum_{J \ni i} N_J \beta_{iJ}, \qquad \gamma_i \equiv \textstyle\sum_{J \ni i} N_J \gamma_{iJ}.$$

Equation (20) implies:

$$\begin{aligned}
\beta_i &= \sum_{J \ni i} N_J \beta_{iJ} \\
&= N_J \alpha \sum_n \omega_{ni} \sum_{J \ni n} \pi_{nJ} \mathbf{1}_{ni}^J \\
&= \frac{\alpha}{\sigma} \sum_n \omega_{ni} \left(\sum_{J \ni n} N_J r_{nJ} \right) \frac{\sum_{J \ni n} N_J r_{nJ} \mathbf{1}_{ni}^J}{\sum_{J \ni n} N_J r_{nJ}} \\
&= \frac{\alpha}{\sigma} \sum_n \omega_{ni} r_n s_{ni},
\end{aligned}$$

where $r_n \equiv \sum_{J' \ni n} N_{J'} r_{nJ'}$ and $s_{ni} \equiv \frac{\sum_J N_J r_{nJ} \mathbf{1}_{ni}^J}{\sum_J N_J r_{nJ}}$. In vector form:

$$\boldsymbol{\beta}' = \frac{\alpha}{\sigma} \boldsymbol{r}' \left(\Omega \circ \boldsymbol{S} \right).$$

Equation (24) implies:

$$\sum_{J \ni i} \frac{d \ln K_n}{d \ln h_{iJ}} = \alpha s_{ni} \omega_{ni} + \delta s_{nq} \omega_{nq} \sum_{J \ni i} \frac{d \ln K_q}{d \ln h_{iJ}}.$$

In vector form:

$$\left[\sum_{J \ni i} \frac{d \ln K_n}{d \ln h_{iJ}} \right] = \alpha \left(I - \delta \Omega \circ \boldsymbol{S} \right)^{-1} \Omega \circ \boldsymbol{S}. \tag{B.1}$$

We thus have:

$$\begin{aligned}
\gamma_i &\equiv \textstyle\sum_{J \ni i} N_J \gamma_{iJ} \\
&= \textstyle\sum_{J \ni i} \dfrac{d \hat{\mathcal{W}}}{d \ln h_{iJ}} \\
&= \frac{\sigma}{\sigma - 1} \left(\frac{\alpha}{\sigma} \sum_n r_n s_{ni} \omega_{ni} + \frac{\delta}{\sigma} \sum_n r_n \sum_m s_{nm} \omega_{nm} \textstyle\sum_{J \ni i} \dfrac{d \ln K_m}{d \ln h_{iJ}} \right).
\end{aligned}$$

In vector form:

$$
\begin{aligned}
\boldsymbol{\gamma}' &= \frac{\sigma}{\sigma-1}\frac{\alpha}{\sigma}\boldsymbol{r}'\left(\Omega\circ\boldsymbol{S}\right)\left(\boldsymbol{I}+\delta\Omega\circ\boldsymbol{S}\left(I-\delta\Omega\circ\boldsymbol{S}\right)^{-1}\right)\\
&= \frac{\sigma}{\sigma-1}\frac{\alpha}{\sigma}\boldsymbol{r}'\left(\Omega\circ\boldsymbol{S}\right)\left(I-\delta\Omega\circ\boldsymbol{S}\right)^{-1}\\
&= \frac{\sigma}{\sigma-1}\boldsymbol{\beta}'\left(I-\delta\Omega\circ\boldsymbol{S}\right)^{-1}.
\end{aligned}
$$

C Extensions

C.1 Heterogeneous Location Size

In our baseline model, the population in each location is normalized to one. We now extend the model to allow for population heterogeneity; precisely, we specify that location i has population μ_i. In this case, the total expenditure in location i is μ_i. The variable profit of variety ν produced in location i that is sold to location n is:

$$\pi_{ni}(\nu) \quad = \quad \mu_n \frac{p_{ni}(\nu)^{1-\sigma}}{\sum_i \int_\nu p_{ni}(\nu)^{1-\sigma} d\nu},$$

and the total variable profit of a plant is:

$$\pi_i(\nu) = b_i \kappa_i(\nu)^{1-\sigma},$$

where b_i is the market access for a plant in location i:

$$b_i \equiv \sum_n \frac{\mu_n}{\sigma} \frac{\tau_{ni}^{1-\sigma}}{\sum_m \int_\nu (\kappa_m(\nu)\tau_{nm})^{1-\sigma} d\nu}. \tag{C.1}$$

Note that, intuitively, the market access b_i for a plant in location i weighs each market n by its population μ_n. In the baseline model, $\mu_n = 1$ (as in (13)).

The utilitarian social welfare is $\mathcal{W} \equiv \sum_n \mu_n U_n$. It is easy to show that Lemma 2 continues to hold in this extended setting, and the marginal impact of a plant's innovation investment on social welfare is equal to the marginal impact on $\hat{\mathcal{W}}$ as defined in (19), with population-adjusted market access terms as in (C.1). Since our subsequent analysis is based on analyzing the impact of innovation investments on $\hat{\mathcal{W}}$, our analysis (including the sufficient statistics) continues to hold in this setting without modification.

C.2 Innovation Decisions along the Extensive Margin

In our baseline model, all J plants of each firm are innovating plants. In this extension, we enrich the model by modeling firms with both innovating and non-innovating plants. Specifically, firms may have heterogeneous efficiencies and make endogenous plant-level innovation decisions along the extensive margin. The set of innovating plants for each firm is thus endogenous. While solving for the equilibrium involves a difficult combinatorial discrete choice problem (cf. Arkolakis, Eckert, and Shi, 2023) and is outside the scope of this paper, we extend our main sufficient statistic result, Proposition 2, to this extended setting. That is, we derive the sufficient statistic for each location's marginal private and social value of innovation in equilibrium.

Formally, we now define a firm $\left(\hat{J}, \varphi\right)$ by its exogenous set of production plants $\hat{J}$ and efficiency φ. All plants produce, but only a subset of plants endogenously innovate. We assume φ is a scalar drawn from a continuous CDF $F_{\hat{J}}$, and that the firm-wide efficiency φ applies to all plants of the firm.

We specify that the marginal cost of a plant in location n is $\kappa_n\left(\hat{J}, \varphi\right) = k_n\left(\hat{J}, \varphi\right)^{\frac{1}{1-\sigma}}$, where:

$$k_n\left(\hat{J}, \varphi\right) = \varphi \prod_{i \in \hat{J}} \max\left\{1, \left(h_i\left(\hat{J}, \varphi\right)^{\alpha} K_i^{\delta}\right)^{\omega_{ni}}\right\}. \tag{C.2}$$

Let $J\left(\hat{J}, \varphi\right) \subseteq \hat{J}$ denote the set of innovating plants of firm $\left(\hat{J}, \varphi\right)$. The productivity of a plant in location n depends on three components:

1. firm-wide efficiency φ;
2. the firm's innovation investments $h_i\left(\hat{J}, \varphi\right)$ in all plant locations $i \in J\left(\hat{J}, \varphi\right)$, including locations $i \neq n$;
3. the location-wide knowledge K_i across all innovating plants in each location $i \in J\left(\hat{J}, \varphi\right)$,

$$K_i \equiv \sum_{\hat{J} \ni i} N_{\hat{J}} \int_{\varphi} k_i\left(\hat{J}, \varphi\right) \mathbf{1}_{i \in J\left(\hat{J}, \varphi\right)} dF_{\hat{J}}\left(\varphi\right). \tag{C.3}$$

Note that, in particular, our specification implies that (i) non-innovating plants do not benefit from local spillovers, nor do they contribute to local knowledge: $k_n\left(\hat{J}, \varphi\right)$ does not benefit from higher K_i, and $k_i\left(\hat{J}, \varphi\right)$ does not contribute to higher K_i, unless $h_i\left(\hat{J}, \varphi\right) > 0$; (ii) non-innovating plants benefit from cross-location spillovers: $k_n\left(\hat{J}, \varphi\right)$ benefits from higher K_i even if $h_n\left(\hat{J}, \varphi\right) = 0$. Both implications are motivated by our empirical evidence (cf. Table 2 and Table 4).

The variable profit of a plant in location n is $\pi_n\left(\hat{J}, \varphi\right) = b_n k_n\left(\hat{J}, \varphi\right)$, where b_n is the market access in location n defined analogously to (13). Given the innovation subsidy $t_{n\hat{J}}\left(\varphi\right)$ to plant n of firm $\left(\hat{J}, \varphi\right)$, the firm solves:

$$\max_{h_n\left(\hat{J}, \varphi\right)} \sum_n \left[b_n \varphi \prod_{i \in \hat{J}} \max\left\{1, \left(h_i\left(\hat{J}, \varphi\right)^{\alpha} K_i^{\delta}\right)^{\omega_{ni}}\right\} - c_n h_n\left(\hat{J}, \varphi\right) / t_n\left(\hat{J}, \varphi\right)\right].$$

In an equilibrium, the private value of a marginal innovation investment in an innovating plant is:

$$\beta_i\left(\hat{J},\varphi\right) \equiv \frac{\partial \sum_{n\in\hat{J}} \pi_n\left(\hat{J},\varphi\right)}{\partial \ln h_i\left(\hat{J},\varphi\right)} = \alpha \sum_{n\in\hat{J}} \pi_n\left(\hat{J},\varphi\right)\omega_{ni}\mathbf{1}_{ni}^{\hat{J}}.$$

Following the same steps as in Lemma 2, we can show that the impact of a marginal innovation investment in an innovating plant $n \in J\left(\hat{J},\varphi\right)$ on social welfare is the same as $\frac{\partial\hat{\mathcal{W}}}{\partial h_n\left(\hat{J},\varphi\right)}|_{\text{holding }\{b_n\}\text{ constant}}$, with $\hat{\mathcal{W}}$ defined as:

$$\begin{aligned}
\hat{\mathcal{W}} &\equiv \sum_n \left(\frac{\sigma}{\sigma-1}\sum_{\hat{J}\ni n} N_{\hat{J}} \int_\varphi \left\{ b_n\varphi \prod_{m\in J\left(\hat{J},\varphi\right)} \left(h_m\left(\hat{J},\varphi\right)^\alpha K_m^\delta\right)^{\omega_{nm}} - \sum_{\hat{J}\ni n} N_{\hat{J}} c_n h_n\left(\hat{J},\varphi\right)\right\} dF_{\hat{J}}\left(\varphi\right)\right) \\
&= \sum_n \left(\frac{\sigma}{\sigma-1}\sum_{\hat{J}\ni n} N_{\hat{J}} \int_\varphi \left\{\pi_n\left(\hat{J},\varphi\right) - \sum_{\hat{J}\ni n} N_{\hat{J}} c_n h_n\left(\hat{J},\varphi\right)\right\} dF_{\hat{J}}\left(\varphi\right)\right).
\end{aligned}$$

Analogous to (22), the social value of a marginal innovation investment in an innovating plant is:

$$\begin{aligned}
& \gamma_i\left(\hat{J},\varphi\right) \\
\equiv\ & N_{\hat{J}}^{-1}\frac{d\hat{\mathcal{W}}}{d\ln h_i\left(\hat{J},\varphi\right)} \\
=\ & \frac{\sigma}{\sigma-1}\frac{\partial \sum_{n\in\hat{J}} \pi_n\left(\hat{J},\varphi\right)}{\partial \ln h_i\left(\hat{J},\varphi\right)} f_{\hat{J}}\left(\varphi\right) \\
& + \frac{\sigma}{\sigma-1} N_{\hat{J}}^{-1} \sum_m \frac{\partial \sum_{\hat{J}'} N_{\hat{J}'} \sum_{n\in\hat{J}'} \int_{\varphi'} \pi_n\left(\hat{J}',\varphi'\right) dF_{\hat{J}'}\left(\varphi'\right)}{\partial \ln K_m} \frac{d\ln K_m}{d\ln h_i\left(\hat{J},\varphi\right)} \\
=\ & \frac{\sigma}{\sigma-1}\alpha \sum_{n\in\hat{J}} \pi_n\left(\hat{J},\varphi\right)\omega_{ni}\mathbf{1}_{ni}^{\hat{J}} f_{\hat{J}}\left(\varphi\right) \\
& + \frac{\sigma}{\sigma-1} N_{\hat{J}}^{-1}\delta \sum_{\hat{J}'} N_{\hat{J}'} \sum_{n\in\hat{J}'} \int_{\varphi'} \sum_{m\in J\left(\hat{J}',\varphi'\right)} \pi_n\left(\hat{J}',\varphi'\right)\omega_{nm}\mathbf{1}_{nm}^{\hat{J}'} dF_{\hat{J}'}\left(\varphi'\right) \frac{d\ln K_m}{d\ln h_i\left(\hat{J},\varphi\right)}. \qquad \text{(C.4)}
\end{aligned}$$

We now derive the private and social value of innovation at the location level. First, we substitute (C.2) into (C.3):

$$K_n \equiv \sum_{\hat{J}\ni n} N_{\hat{J}} \int_{\varphi} \varphi \prod_{i\in\hat{J}} \max\left\{1, \left(h_i\left(\hat{J},\varphi\right)^{\alpha} K_i^{\delta}\right)^{\omega_{ni}}\right\} \mathbf{1}_{n\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi). \tag{C.5}$$

Let

$$s_n\left(\hat{J},\varphi\right) \equiv \frac{N_{\hat{J}}\varphi \prod_{i\in\hat{J}} \max\left\{1, \left(h_i\left(\hat{J},\varphi\right)^{\alpha} K_i^{\delta}\right)^{\omega_{ni}}\right\} \mathbf{1}_{n\in J(\hat{J},\varphi)} f_{\hat{J}}(\varphi)}{\sum_{\hat{J}'\ni n} N_{\hat{J}'} \int_{\varphi'} \varphi' \prod_{i\in\hat{J}'} \max\left\{1, \left(h_i\left(\hat{J}',\varphi'\right)^{\alpha} K_i^{\delta}\right)^{\omega_{ni}}\right\} \mathbf{1}_{n\in J(\hat{J}',\varphi')} dF_{\hat{J}'}(\varphi')}.$$

Differentiating (C.5) with respect to $h_i\left(\hat{J},\varphi\right)$ for a plant with interior innovation investments, we have:

$$\begin{aligned}\frac{d\ln K_n}{d\ln h_i\left(\hat{J},\varphi\right)} &= s_n\left(\hat{J},\varphi\right)\alpha\omega_{ni} \\ &\quad + \int_{\varphi'} \sum_{\hat{J}'} s_n\left(\hat{J}',\varphi'\right) \sum_{q\in J(\hat{J}',\varphi')} \delta\omega_{nq} \frac{d\ln K_q}{d\ln h_i\left(\hat{J},\varphi\right)} F_{\hat{J}'}(\varphi').\end{aligned}$$

Integrating across all innovating plants in location i, we obtain:

$$\begin{aligned}\chi_{ni} &\equiv \sum_{\hat{J}\ni i} \int_{\varphi} \mathbf{1}_{i\in J(\hat{J},\varphi)} \frac{d\ln K_n}{d\ln h_i\left(\hat{J},\varphi\right)} dF_{\hat{J}}(\varphi) \\ &= \left(\sum_{\hat{J}} \int_{\varphi} s_n\left(\hat{J},\varphi\right) \mathbf{1}_{i\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi)\right) \times \alpha\omega_{ni} \\ &\quad + \sum_{q} \left(\sum_{\hat{J}'} \int_{\varphi'} s_n\left(\hat{J}',\varphi'\right) \mathbf{1}_{q\in J(\hat{J}',\varphi')} F_{\hat{J}'}(\varphi')\right) \times \delta\omega_{nq}\chi_{qi}.\end{aligned}$$

Note that $s_{ni} \equiv \sum_{\hat{J}} \int_{\varphi} s_n\left(\hat{J},\varphi\right) \mathbf{1}_{i\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi)$ is the revenue-weighted share of innovating plants in location n that are connected to an innovating plant in location i. Hence:

$$\chi_{ni} = \alpha s_{ni}\omega_{ni} + \delta \sum_{q} s_{nq}\omega_{nq}\chi_{nq}.$$

In matrix form:

$$\boldsymbol{X} \equiv [\chi_{ni}] = \alpha\left(\boldsymbol{I} - \delta\Omega \circ \boldsymbol{S}\right)^{-1} \Omega \circ \boldsymbol{S},$$

which is analogous to equation (B.1) in our baseline model.

The location-specific private value of innovation is:

$$\begin{aligned}
\beta_i &\equiv \sum_{\hat{J}} N_{\hat{J}} \int_{\varphi} \beta_i\left(\hat{J},\varphi\right) \mathbf{1}_{i\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi) \\
&= \sum_{\hat{J}} N_{\hat{J}} \int_{\varphi} \alpha \sum_{n\in\hat{J}} \pi_n\left(\hat{J},\varphi\right) \omega_{ni} \mathbf{1}_{i\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi) \\
&= \frac{\alpha}{\sigma} \sum_n \omega_{ni} \left(\sum_{\hat{J}\ni n} N_{\hat{J}} \int_{\varphi} r_n\left(\hat{J},\varphi\right) \mathbf{1}_{i\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi) \right) \\
&= \frac{\alpha}{\sigma} \sum_n \omega_{ni} \hat{s}_{ni} r_n,
\end{aligned}$$

where r_n is the total revenue of all plants in location n, and $\hat{s}_{ni}$ is the revenue-weighted fraction of all plants in location n that are connected to an innovating plant in location i. Note the distinction between $\hat{s}_{ni}$ and s_{ni}: the former refers to all plants in location n; the latter refers to all innovating plants in location n. In vector form:

$$\boldsymbol{\beta}' = \frac{\alpha}{\sigma} \boldsymbol{r}' \left(\boldsymbol{\Omega} \circ \hat{\boldsymbol{S}} \right).$$

Finally, to derive the location-specific social value of innovation, we integrate (C.4) across all innovating plants in location i:

$$\begin{aligned}
\gamma_i &\equiv \sum_{\hat{J}} N_{\hat{J}} \int_{\varphi} \gamma_i\left(\hat{J},\varphi\right) \mathbf{1}_{i\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi) \\
&= \frac{\sigma}{\sigma-1} \alpha \sum_{\hat{J}} N_{\hat{J}} \int_{\varphi} \left(\sum_{n\in\hat{J}} \pi_n\left(\hat{J},\varphi\right) \omega_{ni} \mathbf{1}_{ni}^{\hat{J}} f_{\hat{J}}(\varphi) \right) \mathbf{1}_{i\in J(\hat{J},\varphi)} dF_{\hat{J}}(\varphi) \\
&\quad + \frac{\sigma}{\sigma-1} \delta \left(\sum_{\hat{J}'} N_{\hat{J}'} \sum_{n\in\hat{J}'} \int_{\varphi'} \sum_m \pi_n\left(\hat{J}',\varphi'\right) \omega_{nm} \mathbf{1}_{nm}^{\hat{J}'} dF_{\hat{J}'}(\varphi') \chi_{mi} \right) \\
&= \frac{1}{\sigma-1} \left(\alpha \sum_n r_n \hat{s}_{ni} \omega_{ni} + \delta \sum_n r_n \sum_m s_{nm} \omega_{nm} \chi_{mi} \right).
\end{aligned}$$

In vector form:

$$
\begin{aligned}
\boldsymbol{\gamma}' &= \frac{\alpha}{\sigma-1}\left(\boldsymbol{r}'\left(\Omega \circ \hat{\boldsymbol{S}}\right) + \delta \boldsymbol{r}'\left(\Omega \circ \hat{\boldsymbol{S}}\right)\left(\boldsymbol{I} - \delta\Omega \circ \boldsymbol{S}\right)^{-1}\Omega \circ \boldsymbol{S}\right) \\
&= \frac{\alpha}{\sigma-1}\boldsymbol{r}'\left(\Omega \circ \hat{\boldsymbol{S}}\right)\left(\boldsymbol{I} - \delta\Omega \circ \boldsymbol{S}\right)^{-1} \\
&= \frac{\sigma}{\sigma-1}\boldsymbol{\beta}'\left(\boldsymbol{I} - \delta\Omega \circ \boldsymbol{S}\right)^{-1}.
\end{aligned}
$$

We summarize these results in the following proposition.

Proposition 3. *With productivity heterogeneity and endogenous innovation investments along the extensive margin, the location-specific private value of innovation investments is* $\boldsymbol{\beta}' = \frac{\alpha}{\sigma}\boldsymbol{r}'\left(\Omega \circ \hat{\boldsymbol{S}}\right)$*, where* $\hat{\boldsymbol{S}}$ *is the matrix whose* ni*-th entry captures the revenue-weighted share of all plants in location* n *that are connected to an innovating plant in location* i*. The corresponding social value of innovation investments is* $\boldsymbol{\gamma}' = \frac{\sigma}{\sigma-1}\boldsymbol{\beta}'\left(\boldsymbol{I} - \delta\Omega \circ \boldsymbol{S}\right)^{-1}$.