# The Misallocated Innovation Subsidies in Space

Sheng Cai, Ningyuan Jia, Weizhe Wang, Wei Xiang *

September 16, 2026

**Abstract**

Innovation subsidies can lower productivity when they draw scarce researchers away from adopting external ideas. We study this trade-off in a spatial model in which the local supply of research labor is inelastic in the medium run, researchers choose between frontier innovation and technology adoption, and trade determines the pool of ideas available for adoption. The model distinguishes two spatial forces: the local cost of an innovation subsidy rises with accessible knowledge, while its external benefit rises with the locations reach in the knowledge-diffusion network. Using 3.25 million Chinese manufacturing firm-year observations from 2000-2013, we find that subsidy receipt increases own innovation but crowds out imitation. This crowd-out, and the resulting productivity loss, becomes larger with external knowledge access. Instrumenting knowledge access with high-speed rail expansion along historical rail corridors and foreign import-supply shocks yields the same gradient. We use these estimates and inter-provincial trade flows to quantify the model across thirty provinces. The calibrated innovation wedges vary more than twentyfold, and a 10 percent increase in the wedge raises the local technology frontier in fifteen provinces but lowers it in fifteen. Welfare-optimal implicit subsidy rates range from 5 to 33 percent and tend to be higher in provinces closer to the world frontier. The external benefit of a province's innovation instead depends on how widely its ideas spread to other provinces through the trade network. Replacing the calibrated wedges with this schedule raises manufacturing TFP by 11.26 percent and worker welfare by 4.25 percent.

*We thank Ezra Oberfield, Zebang Xu for helpful comments. Sheng Cai: City University of Hongkong, shengcai@cityu.edu.hk; Ningyuan Jia: UIBE, jia_ningyuan@uibe.edu.cn; Weizhe (Aaron) Wang: Cornell University, ww572@cornell.edu; Wei Xiang: University of Michigan, xwe@umich.edu.

# 1 Introduction

The standard case for innovation subsidies starts from a knowledge externality. Firms cannot capture the full social value of the ideas they create, so private research incentives may be too weak (Nelson, 1959; Arrow, 1962; Bloom, Schankerman and Van Reenen, 2013). This argument usually treats research as a single activity and asks how much firms should conduct. Yet firms also devote research effort to imitation, by which we mean identifying, adopting, and adapting ideas developed elsewhere (Cohen and Levinthal, 1989; Griffith, Redding and Van Reenen, 2004). When the local supply of research labor is inelastic in the medium run, a subsidy to frontier innovation can therefore draw researchers away from imitation. Whether the subsidy raises productivity then depends not only on the ideas it creates, but also on the external ideas that the location no longer absorbs.

This paper studies how this trade-off changes across space. Geography matters through two distinct margins. A location's *knowledge access* determines its pool of imitable ideas and hence the local cost of redirecting researchers from imitation toward innovation. Its *diffusion reach* determines how widely its new ideas benefit other locations and hence the external return to local innovation. These margins need not coincide. A location may have access to a rich stock of outside knowledge without occupying a central position in the diffusion network. The optimal innovation incentive must balance the local productivity loss from displaced imitation against the external gain from creating ideas that diffuse elsewhere.

China provides a useful setting for studying this balance. Innovation and R&D subsidies to manufacturing firms rose from RMB 31.6 billion in 2000 to RMB 140.9 billion in 2013, and roughly one in seven firms received a subsidy in a typical year. The policy also varied sharply across space. Figure 1 shows that subsidies reached the largest share of firms in many inland and northwestern cities, which had relatively weak trade links and limited access to outside knowledge. By contrast, the largest numbers of subsidized firms were concentrated in major northeastern and coastal cities, where the subsidy rate per firm was lower. This spatial pattern could reflect efficient targeting: the cost of displaced imitation may be lower where external knowledge is scarce. It could also reflect policy distortions that bear little relation to the social value of innovation. We ask which interpretation is consistent with the data and what allocation of innovation incentives would maximize welfare.

We begin with evidence that innovation policy changes the composition of research. At the city level, we combine the patent records of 247 prefectural cities from 2000 to 2013 with a measure of local industrial policy based on the accumulated stock of policy documents. We classify patents as breakthrough or imitative following Liu and Wang (2025). A one-log-unit increase in the policy stock raises the breakthrough-patent rate by 3.3 percentage points and lowers the imitative-patent rate by 2.0 percentage points. Thus, policy changes not only how much cities patent, but also the type of knowledge they produce.

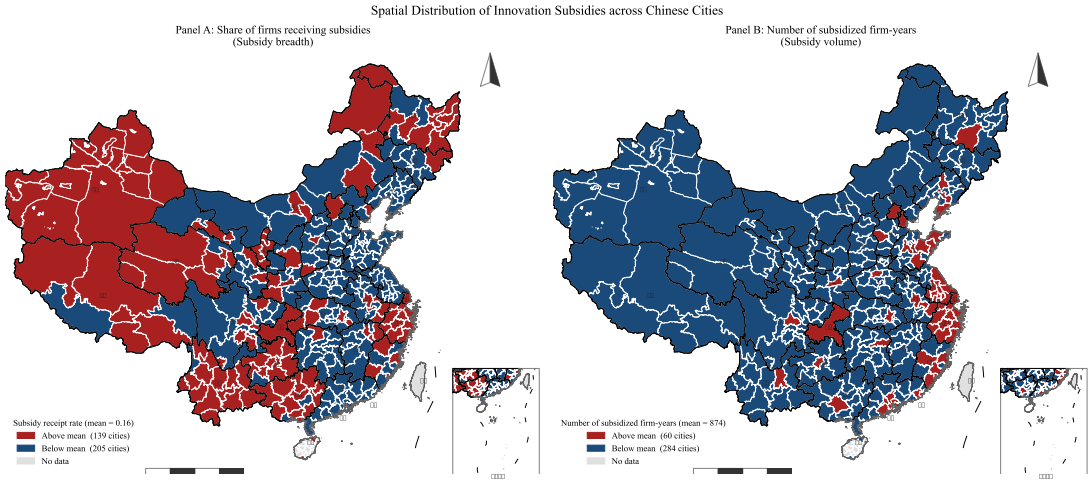



Figure 1: Spatial distribution of innovation subsidies across Chinese cities

*Notes:* Panel A plots the share of firm-year observations receiving government innovation subsidies in each prefectural city (subsidy breadth). Panel B plots the total number of subsidized firm-year observations (subsidy volume). Both panels use the firm-year analysis sample with non-missing TFP and subsidy status. Red (blue) cities lie above (below) the cross-city mean. Both panels show where subsidies flow, not how well they work.

We next measure imitation at the firm level. Firms report their own patents, but they do not report the technologies they absorb from others. Following Cai et al. (2025), we construct an imitation index from the part of firm productivity that cannot be explained by the firm's own patenting, observable inputs, or common industry shocks. We validate this index using patent data that do not enter its construction. Cities with a higher imitation index cite outside patents more intensively and incorporate a larger and more similar share of technologies previously developed elsewhere. In addition, a text-based classification of imitative patents declines with the accumulation of local innovation policy. These independent measures point to the same shift from imitation toward own innovation.

Our main empirical analysis uses 3.25 million firm-year observations between 2000 and 2013. We estimate an interacted difference-in-differences specification that compares firms before and after subsidy receipt and allows the effect to vary with local knowledge access. Subsidies raise own innovation and reduce imitation on average. The main result is spatial: both responses become stronger as access to outside knowledge rises, and the imitation response eventually dominates. Subsidies therefore raise firm productivity across most of the knowledge-access distribution, and the net effect declines with access, turning negative only among the most knowledge-rich cities. We obtain the same gradient when we measure access with transport costs: innovation subsidies are less harmful in isolated locations and more harmful in well-connected ones.[1]

The interaction raises a separate identification concern. Knowledge access may be correlated with local growth prospects or with the endogenous placement of subsidies. We address this concern using two sources of variation that shift access to outside knowledge. The first uses high-speed rail expansion along routes predicted by the historical slow-rail network. A city's exposure therefore depends on its position in a network designed decades before the subsidy outcomes we study.

[1] Cai et al. (2025) establish the average causal effect of subsidy receipt on the innovation–imitation margin using three sources of policy variation. This paper studies a different object: how that effect varies with the external knowledge available to the firm.

The second combines a city's predetermined import composition with subsequent productivity growth among foreign suppliers. This measure shifts the external knowledge embodied in a city's imports through supply changes abroad rather than local demand. Both strategies yield a negative subsidy-by-access effect on imitation and productivity and a positive effect on own innovation. The estimated gradients are larger than their ordinary-least-squares counterparts, consistent with attenuation from measurement error in constructed knowledge access.

To interpret these findings and measure their aggregate consequences, we develop a quantitative spatial model of research allocation and trade-mediated knowledge diffusion. Each location has a stock of researchers who choose between frontier innovation and imitation. Innovation creates ideas that can diffuse to other locations. Imitation combines local research effort with a pool of foreign ideas accumulated through bilateral trade. A policy wedge raises the private payoff to innovation and shifts researchers away from imitation. This shift is more costly where the imitable pool is large. At the same time, innovation creates an external benefit by enlarging the pools available to other locations. This benefit is larger for locations with greater diffusion reach.

We embed this mechanism in an Eaton and Kortum (2002) trade environment with correlated productivity draws as in Lind and Ramondo (2023). The resulting structure links trade, idea diffusion, and the evolution of each location's technology frontier. It also delivers a transparent policy rule. Starting from the locally efficient research allocation, a small innovation incentive creates a second-order local loss but a first-order diffusion gain. The optimal incentive remains positive because a location does not internalize the value of its innovations to others. Its level, however, depends on the balance between the local imitation cost, governed by knowledge access, and the external diffusion benefit, governed by network reach. Distance from the world frontier alone is therefore not a sufficient guide for place-based innovation policy.

We quantify the model for thirty Chinese provinces and an aggregate rest of the world. The calibration combines the reduced-form innovation and imitation responses with provincial trade flows, migration flows, researcher stocks, and observed productivity. The relative innovation and imitation responses identify how subsidies reallocate research. The instrumental-variable response of imitation to knowledge access identifies the learning elasticity. Provincial trade flows discipline the pools of ideas that locations can access and the reach of the ideas they create. The remaining trade, productivity, and amenity fundamentals are recovered from equilibrium conditions.

The model recovers a province-specific innovation-payoff wedge that measures how strongly the local policy environment tilts research toward innovation relative to imitation. The calibrated wedges range from 0.25 to 6.59, a difference of more than twentyfold. A 10 percent increase in a province's wedge raises its technology frontier in fifteen provinces and lowers it in the other fifteen. The negative effects occur where current incentives have already pushed research beyond the point at which additional innovation compensates for foregone imitation. Thus, the same policy can correct an innovation externality in one province and deepen a research-allocation distortion in another.

We then solve for the province-specific incentive schedule that maximizes worker welfare. To isolate the research-allocation channel, the baseline exercise treats the policy wedge as an implicit subsidy without a fiscal resource cost. The welfare-optimal implicit subsidy rates range from 5 to 33 percent, compared with the much wider dispersion in the calibrated economy. Welfare-optimal rates tend to be higher in provinces closer to the technology frontier, where the local cost of displaced imitation is smaller. The planner also values diffusion reach because innovations from well-connected provinces spread widely. Replacing the calibrated wedges with the optimal schedule

raises manufacturing TFP by 11.26 percent and worker welfare by 4.25 percent. Compressing extreme wedges accounts for most of the gain. The optimal schedule retains moderate spatial variation.

Our first contribution is to the literature on innovation and place-based subsidies. Empirical work estimates how subsidies affect innovation, employment, and productivity at treated firms or locations (Howell, 2017; Kline and Moretti, 2014; Busso, Gregory and Kline, 2013), while quantitative spatial models study optimal place-based transfers in the presence of firm sorting, agglomeration, and worker mobility (Gaubert, 2018; Fajgelbaum and Gaubert, 2020). This literature typically evaluates a subsidy through the market failure at the targeted location. We identify an additional margin: innovation policy changes how researchers use their time. Because the value of the displaced activity depends on external knowledge access, a subsidy that benefits a recipient can reduce productivity in the targeted location and alter the knowledge available elsewhere.

Our second contribution is to the literature on innovation, imitation, and technology adoption. Models of appropriate growth policy emphasize distance from the frontier (Acemoglu, Aghion and Zilibotti, 2006; Howitt and Mayer-Foulkes, 2005). Other work models adoption as an active choice through search, learning, or investment (Perla and Tonetti, 2014; Lucas and Moll, 2014; Benhabib, Perla and Tonetti, 2014, 2021; König, Lorenz and Zilibotti, 2016). Closest to our research-allocation margin, König et al. (2022) structurally estimate firms' choices between innovation and imitation in China and Taiwan. We add a policy wedge and a spatially varying pool of imitable ideas. We then use policy-induced variation to identify how the cost of shifting research toward innovation depends on access to that pool.

Our third contribution is to work on knowledge diffusion through trade and production networks. Trade changes not only market access but also the set of ideas that firms encounter (Buera and Oberfield, 2020; Sampson, 2016; Perla, Tonetti and Waugh, 2021). Research on knowledge spillovers measures these links using geographic distance, patent citations, and technological proximity (Jaffe, Trajtenberg and Henderson, 1993; Moretti, 2004; Bloom, Schankerman and Van Reenen, 2013). We make both sides of the knowledge network endogenous: trade determines the ideas a location can imitate, while local innovation changes the ideas available to its partners. This distinction between inward knowledge access and outward diffusion reach is what connects the reduced-form gradient to the optimal spatial policy.

Finally, the paper contributes to quantitative work on spatial and factor misallocation. Existing research measures distortions in static factor use or follows subsidies through production networks (Hsieh and Klenow, 2009; Hsieh et al., 2019; Liu, 2019). We instead recover a dynamic wedge on the allocation of a fixed local research stock. Embedding this wedge in spatial general equilibrium allows us to measure both the local opportunity cost of displaced imitation and the external value of the innovations that diffuse through the trade network.

Section 2 presents the data and measurement, and Section 2.2.2 the empirical results. Section 4 develops the model. Section 5 describes the calibration. Section 6 presents the counterfactuals, and Section 7 concludes.

# 2 Empirical Evidence

## 2.1 Data

Our empirical analysis draws on four complementary data sources. Appendix A provides additional construction details and a summary statistics table.

**Firm-level production data.** The backbone of our analysis is the *Annual Survey of Industrial Enterprises in China* (ASIEC), compiled by the National Bureau of Statistics. The survey constitutes a near-census of all above-scale industrial enterprises (those with annual revenues exceeding RMB 5 million), spanning 1998 to 2013. For each firm-year, the data report total output, value-added, employment, the gross book value of fixed assets, intermediate input expenditure, and government subsidy income recorded in the income statement. This last item allows us to observe at the firm level not only whether a subsidy was received but also its magnitude, making it possible to construct event-study samples tied to the first year of subsidy receipt. After imposing a minimum two-year panel requirement and dropping observations with implausible factor ratios, the sample contains approximately 3.5 million firm-year observations.

**Patent records.** We obtain the universe of Chinese patent applications from the *China National Intellectual Property Administration* (CNIPA), encompassing all invention, utility-model, and design patents applied for between 1985 and 2018. The data record filing dates, IPC technology classifications, legal status, and forward and backward citation lists. We match patents to ASIEC firms using a probabilistic name–address algorithm supplemented by enterprise registration codes, successfully linking approximately 2.8 million firm-year observations to at least one patent filing. We use these records to construct measures of patenting volume, technology-class diversity, patent quality, and interregional citation flows.

**Industrial policy documents.** To identify subsidy-granting events, we collected the full text of all subsidy-related policy documents (project approvals, selection announcements, and disbursement notices) issued by municipal, provincial, and central-government bodies between 2000 and 2015. From these documents we extract the subsidising authority, the stated policy objective, and the grant amount where disclosed. The documents allow us to classify subsidies by stated purpose.

**City-level yearbooks and trade data.** We supplement the firm-level panel with city-year data from provincial and municipal statistical yearbooks, covering GDP, fiscal activity, population, infrastructure, and higher-education capacity. To construct the import-based knowledge instrument (Section 2.2.6), we use customs transaction records from China's General Administration of Customs, disaggregated to the HS-6 product–source-country level, following the supply-push shift-share design of Autor, Dorn, and Hanson (2013). High-speed rail data (route-kilometres in operation by city and year) are compiled from Ministry of Transport annual reports.
*Sample.* The baseline sample spans 2000–2013 and covers manufacturing firms in prefectural-level cities with non-missing TFP estimates. The stacked panel contains 3,246,795 firm-year observations across 371 prefectural cities before imposing outcome-specific regression restrictions; the TFP summary-statistics sample contains 2,617,487 observations.

## 2.2 Measurements, Facts, and Validations

This section introduces our three core outcome measures (TFP, technology-oriented innovation, and imitation) and the city-level spatial variables that characterise the local knowledge environment. We then present a series of reduced-form validations that provide the empirical basis for the spatial theory in Section 4.

**Total Factor Productivity.** We measure firm-level TFP following Ackerberg, Caves and Frazer (2015). Assuming a Cobb–Douglas production function,

$$y_{ict} = \beta_k k_{ict} + \beta_l l_{ict} + \beta_m m_{ict} + \omega_{ict} + \varepsilon_{ict}, \tag{1}$$

where $y_{ict}$ is log output, $k$, $l$, $m$ denote log capital, labour, and intermediate inputs, $\omega_{ict}$ is the latent TFP component, and $\varepsilon_{ict}$ captures idiosyncratic shocks. We recover $\omega_{ict}$ using the two-stage proxy-variable estimator, which exploits the invertibility of the intermediate-input demand function and assumes TFP follows a first-order Markov process, $\omega_{ict} = g(\omega_{ict-1}) + \xi_{ict}$. Appendix D describes the estimation procedure and reports robustness to alternative estimators.

**Technology Innovation.** We measure firm-level innovation using two complementary indicators. The first is the count of invention-type patent applications. The second is a composite technology-innovation intensity index:

$$\widetilde{Q}_{ict} = \begin{cases} \ln\left(Z_{ict}^{\text{All}} \cdot \mathit{PatentKnowledge}_{ict}^{\text{All}}\right), & Z_{ict}^{\text{All}}\,\mathit{PatentKnowledge}_{ict}^{\text{All}} > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathit{PatentKnowledge}_{ict}^{\text{All}} = 1 - \sum_g (Z_{icgt}^{IU}/Z_{ict}^{\text{All}})^2$ is a Herfindahl-based diversity index across IPC technology groups, $Z_{icgt}^{IU}$ denotes invention and utility-model patents in group $g$, and $Z_{ict}^{\text{All}}$ is total patents filed.[2]

**Imitation.** Since imitation is not directly observable, we construct a residual-based measure that isolates productivity gains attributable to learning from external sources rather than own innovation. We project firm-level TFP onto innovation intensity and a comprehensive set of controls while absorbing firm and year fixed effects:

$$\widehat{\omega}_{ict} = \alpha\,\widetilde{Q}_{ict} + \Gamma X_{ict} + \mu_i + \lambda_t + \varepsilon_{ict}, \quad \mathit{Immi}_{ict} \equiv \widehat{\varepsilon}_{ict}. \tag{3}$$

The baseline controls include firm age, firm history, export status, and predetermined city trends in population, GDP per capita, and industrial structure. This residual captures productivity variation not explained by firms' own patenting or observable inputs. Firms in knowledge-rich environments face larger accessible pools of technologies to absorb, so $\mathit{Immi}_{ict}$ should be systematically higher in cities with greater knowledge accessibility, a prediction we validate below.

[2] Multiplying the diversity index by total patenting volume scales the measure by patenting intensity; the logarithmic transformation mitigates skewness. The zero assignment retains firms for which the logarithm is undefined. We interpret $\widetilde{Q}_{ict}$ as technology-orientation intensity rather than a quality index. See Appendix D for construction details.

The measure's properties are documented in a companion paper (Cai et al., 2025), which establishes four results on the same firm-patent-subsidy sample. First, the residual is serially persistent: its first-order autocorrelation is 0.179, against a mechanical benchmark of $-0.215$ under within-transformation, giving a corrected value of 0.395. Second, $Immi_{ict}$ responds to the external knowledge environment: a leave-one-out city-mean productivity index raises the residual by 0.366, significant at the 1 percent level, confirming the measure tracks conditions outside the firm rather than firm-specific noise. Third, imitative patenting at $t$ predicts *Immi* at $t+1$ and $t+2$ (coefficients of 0.039 and 0.038, both significant) but not at $t$ or earlier, consistent with the lag between knowledge absorption and its realization in measured output. Fourth, subsidies reduce the forecastable component of *Immi* by 0.211 standard deviations one year after receipt, against $-0.027$ for the surprise component, which rules out government selection on firms in transitorily good years as the explanation for the measured decline. We carry these properties into the present analysis and report only the spatial extension of the second result: how the measure's responsiveness to outside knowledge varies across cities with different accessibility levels. The imitative-patent rate in Section 3.1, constructed from patent-text cosine similarity in the CNIPA corpus and sharing no variable with $Immi_{ict}$, provides a cross-check from an entirely different source: it falls systematically as local IP accumulation rises (Table 2). The city-level patent-composition shift and the firm-level productivity residual thus trace the same reallocation of research effort from two independent bodies of data.

**Spatial knowledge accessibility and iceberg costs.** Two city-level variables characterise the external knowledge environment. *Accessible Knowledge* aggregates, for each city $c$ in year $t$, the productivity stocks of nearby outside cities, discounted by road travel cost:

$$A_{ct} = \ln\left(1 + \sum_{d \in \mathcal{N}_{30}(c,t)} \frac{\exp(\bar{\omega}_{dt})}{r_{cdt}}\right), \tag{4}$$

where $\bar{\omega}_{dt}$ is the city-average industry-adjusted TFP index in city $d$ and year $t$, $r_{cdt}$ is bilateral road travel cost, and $\mathcal{N}_{30}(c,t)$ is the set of the thirty lowest-cost outside cities from $c$ in year $t$. To remove industry-composition bias, $\bar{\omega}_{dt}$ is constructed as the city-year mean of firm-level TFP z-scored within narrow (small-category) industry$\times$year cells, so that $\bar{\omega}_{dt}$ reflects within-industry productivity standing rather than the industry mix of the city. *Iceberg Cost* is the corresponding road-cost proxy, defined as the scaled sum of costs to those nearby outside cities. Appendix Figure 14 maps both variables in 2013 and illustrates their substantial East–West dispersion.

### 2.2.1 Industrial Policy in Space

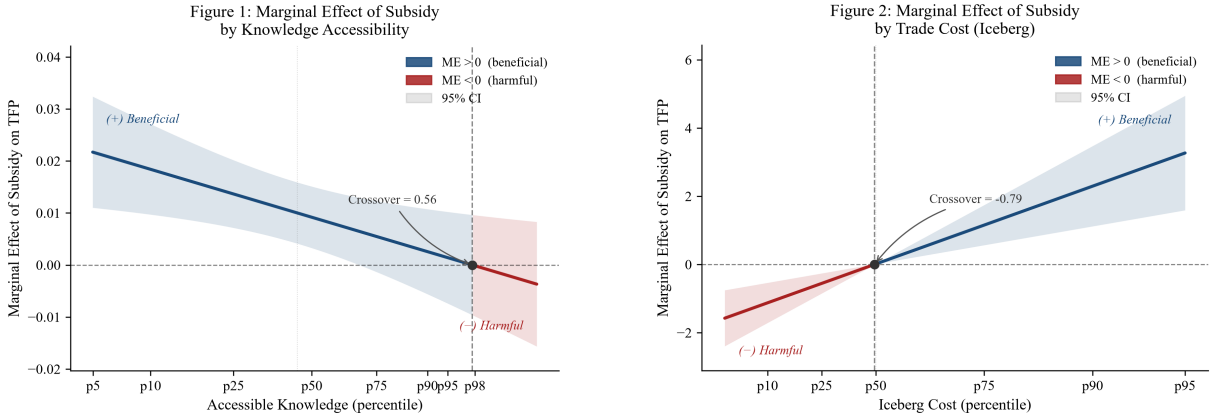
Three cross-city facts motivate the spatial model before we turn to regression evidence. Table 11 (Appendix B) ranks all 244 qualifying prefectural cities on four dimensions: subsidy receipt rate, average firm TFP, technology innovation intensity, and our imitation measure. The pattern is consistent across the distribution. Among the ten cities with the highest subsidy rates, average TFP ranks 166th out of 244 (lower third of the productivity distribution), average technology innovation ranks 71st (upper third), and average imitation ranks 144th (below median). The ten cities with the lowest subsidy rates show the mirror image: average TFP ranks 53rd (top quarter), average innovation ranks 187th (bottom quarter), and average imitation ranks 105th (above median). The

contrast survives within broad industry groups and is present in both coastal and inland subsamples, so city size and industrial composition do not explain it. The raw ranking is consistent with the crowd-out mechanism: subsidies flow to cities where firms innovate intensively but absorb little external knowledge, while the cities that learn most from their neighbors are precisely those that receive the fewest subsidies.

Table 13 (Appendix B) documents the geography of industrial policy and its relationship with local knowledge conditions. Columns (1)–(2) establish the baseline: both city-level R&D subsidy totals and individual firm subsidies are positively associated with average firm TFP, consistent with the broad view that place-based industrial policies correlate with productivity. A one-standard-deviation increase in city-level subsidy spending is associated with a 0.7 percent rise in average firm TFP, robust to city and year fixed effects and a rich set of city controls.

The more revealing patterns emerge from the interaction terms. Columns (3)–(4) show that the positive subsidy–TFP association *diminishes sharply* as accessible knowledge rises: the coefficient on R&D Subsidy City Total × Accessible Knowledge is −0.018, significant at the 1 percent level, so the average relationship reverses sign in the upper quartile of the knowledge distribution. Columns (5)–(6) reveal the symmetric pattern for trade costs: the subsidy–TFP gradient increases with iceberg costs (+0.019, significant at the 1 percent level), meaning the fiscal returns to subsidisation are largest precisely where firms are most isolated from external knowledge. Together, these facts motivate the spatial model in Section 4: innovation subsidies are most productive where imitation is least feasible.

Figure 2 translates these two interaction effects into marginal-effect plots that show the full distribution of the policy's productivity impact across cities. Each panel plots the estimated marginal effect of subsidy receipt on average firm TFP, together with a 95 percent confidence band constructed by the delta method, as a function of the moderating spatial variable. Panel A uses accessible knowledge as the horizontal axis. The marginal effect follows the linear function $\widehat{\mathrm{ME}} = 0.010 - 0.018 \times K_{ct}$, where $K_{ct}$ is the city's demeaned accessible knowledge index. The effect is positive (blue) across most of the knowledge distribution and turns negative (red) only at the crossover point $K \approx 0.556$, which falls between the 97th and 98th percentilesapproximately the top 2 percent of the sample. Panel B uses the demeaned road-cost proxy: $\widehat{\mathrm{ME}} = 0.015 + 0.019 \times C_{ct}$. The effect is increasing in transport costs, so subsidies add more productivity precisely where firms are most geographically isolated from external knowledge sources. Both panels display the inland–coastal divide in policy effectiveness that Figure 3 maps geographically: the coastal cities that sit at the right tail of the knowledge distribution and the left tail of transport costs are the cities most damaged by the current allocation of subsidies.




A. By accessible knowledge

B. By iceberg trade cost

Figure 2: Marginal effect of subsidy receipt on firm TFP, by local knowledge environment

*Notes:* Each panel plots the estimated marginal effect of subsidy receipt on firm TFP as a function of the city's spatial characteristic, with a 95% confidence band (shaded, delta method). Panel A: $\widehat{\text{ME}} = \hat{\beta}_1 + \hat{\beta}_2 K_{ct}$ with $(\hat{\beta}_1, \hat{\beta}_2) = (0.010, -0.018)$ from Table 13, col. (3) (Appendix B). Panel B: $\widehat{\text{ME}} = \hat{\beta}_1 + \hat{\beta}_2 \tau_{ct}$ with $(0.015, +0.019)$ from col. (5). Blue (red) segments indicate positive (negative) marginal effects. The crossover in Panel A at $K \approx 0.556$ falls between the 97th and 98th percentiles; only the top $\sim$2% of cities face a negative marginal effect. Horizontal axis shows percentiles of each variable.

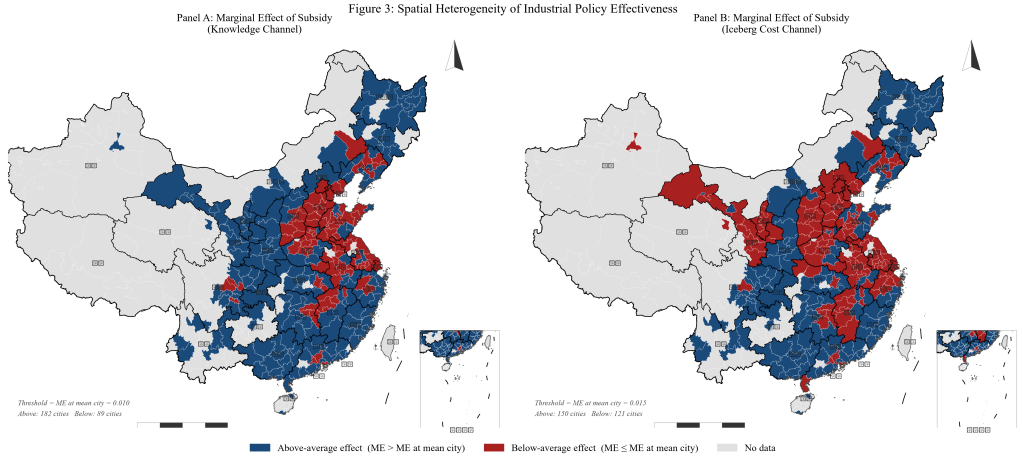



Figure 3: Spatial heterogeneity in the marginal productivity effect of innovation subsidies

*Notes:* Each city is colored by whether its estimated marginal TFP effect of subsidy receipt lies above (blue) or below (red) the mean-city effect. Panel A uses the knowledge-accessibility channel: the city-specific marginal effect is $\widehat{\text{ME}} = 0.010 - 0.018 \times K_{ct}$, where $K_{ct}$ is accessible knowledge from Table 13, column (3); threshold = mean-city marginal effect of 0.010. Panel B uses the iceberg-cost channel: $\widehat{\text{ME}} = 0.015 + 0.019 \times \tau_{ct}$ from column (5); threshold = 0.015. Blue cities experience above-average subsidy effectiveness; red cities below-average. Both panels display a pronounced inland–coastal divide, consistent with Figures 2A–B.

Table 1: Knowledge Accessibility, Trade Costs, and Heterogeneous Innovation

| | *Panel: Universe of Industrial Firms* | | | |
|---|---|---|---|---|
| | Tech Innovation | | Imitation | |
| | (1) | (2) | (3) | (4) |
| Iceberg Cost | 0.020*** | | −0.019** | |
| | (0.003) | | (0.008) | |
| Accessible Knowledge | | −0.067*** | | 0.290*** |
| | | (0.004) | | (0.010) |
| Observations | 2,804,749 | 2,804,749 | 2,616,207 | 2,616,207 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| City FE | ✓ | ✓ | ✓ | ✓ |
| City Controls | ✓ | ✓ | ✓ | ✓ |

*Notes:* The table reports OLS estimates of the relationship between city-level knowledge accessibility and firm-level innovation outcomes. Tech Innovation is the composite technology-orientation index $\widetilde{Q}_{ict}$; Imitation is the residual-based measure $Immi_{ict}$. Iceberg Cost is the road-cost proxy described in Appendix D; Accessible Knowledge aggregates outside-city industry-adjusted TFP discounted by road travel cost (Appendix D). Standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% levels.

### 2.2.2 Knowledge Diffusion in Space

Table 1 establishes two first-order facts about the spatial determinants of innovation and imitation. The dependent variables are the technology-innovation index $\widetilde{Q}_{ict}$ (columns 1–2) and the imitation residual $Immi_{ict}$ (columns 3–4). The sign pattern is consistent throughout. Higher road-cost exposure is associated with *more* own-innovation (coefficient 0.020, significant at the 1 percent level) but *less* imitation (coefficient −0.019, significant at the 5 percent level): firms isolated from external knowledge pools are pushed toward in-house R&D. Conversely, greater accessible knowledge *reduces* a firm's own-innovation intensity (coefficient −0.067, significant at the 1 percent level) while substantially raising imitation (coefficient 0.290, significant at the 1 percent level): spillovers from nearby productive cities crowd out independent R&D and enable external technology absorption. These patterns replicate the core trade-off of the model: innovation and imitation are substitutes, with the spatial knowledge environment as the primary margin.

To validate that $Immi_{ict}$ captures genuine inter-firm knowledge flows, Appendix Table 14 relates city-year imitation to the CNIPA patent network along two channels that share no data with the productivity residual.

The first is the citation record. Column (1) regresses the citations a city's patents make to another city's patents, at a five-year horizon and divided by the citing city's own patent count, on the city's imitation index. The coefficient is 0.005, significant at the 1 percent level. Entering the index of the city being cited alongside it leaves that second coefficient at −0.001 and indistinguishable from zero, so the relation runs through the side doing the absorbing. Scaling by the citing city's patent count is what separates the two: on the raw citation count both cities load positively, because a city that files more patents both cites and is cited more.

The second channel is the text of the patents. Columns (2)–(3) show that a city scoring high

on the imitation index takes up a larger share of technology already developed elsewhere, and takes it up more closely: the share of city-pair patent comparisons clearing the 90th-percentile similarity threshold rises by 0.4 percentage points, and the mean similarity among those comparisons rises by 0.3 log points, both significant at the 1 percent level.

All three coefficients survive on an imitation index built without filling gaps in a firm's series, keeping their sign, their significance and their order of magnitude, so the validation does not rest on how the index is constructed. All three outcomes are also free of patent volume. The raw count of similar patents is not, and Appendix D records it separately.

# 3 Main Results

## 3.1 Macro Evidence: IP and Innovation Dynamics

We begin with city-level evidence on how the accumulation of IP documents shapes the composition of the local patent portfolio. Using the full CNIPA corpus matched to 247 prefecture-level cities over 2000–2013, we ask whether cities that built up larger stocks of IP documents experienced a systematic reallocation from imitative toward breakthrough innovation.

**Patent classification.** We classify each patent along two quality dimensions following Liu and Wang (2025). A patent is designated as *breakthrough* (BK) if its innovation quality score falls in the top decile of its technology class–year distribution (using a China-domestic threshold). The score is the log ratio of the patent's average forward cosine similarity (to all patents filed in the subsequent five years) to its average backward cosine similarity (to all patents filed in the preceding five years) (Kelly et al., 2021). A high score indicates that the patent reoriented the direction of subsequent research rather than extending prior art. A patent is designated as *imitative* (Mimic) if its maximum cosine similarity to any patent filed in the preceding five years (within the same technology class) falls in the top decile; this identifies patents that closely reproduce recent prior art.

The two labels are extremes of the quality distribution, not a partition of it. Each captures roughly one patent in ten by construction, so the majority of patents are neither breakthrough nor imitative. This matters for how the city-year outcomes should be read, and we return to it when defining them below. Appendix I describes the construction in detail; Appendix Table 12 reports summary statistics.

**Specification.** To isolate the within-city, over-time relationship between IP accumulation and innovation composition, we estimate

$$\Delta y_{ct} = \beta \, \Delta \ln(\text{CumIP}_{ct} + 1) + \gamma_t + \varepsilon_{ct}, \tag{5}$$

where $y_{ct}$ denotes one of three outcomes for city $c$ in year $t$: (i) the *imitative-patent rate* $\equiv n_{ct}^{\text{Mimic}}/n_{ct}^{\text{total}}$; (ii) the *breakthrough-patent rate* $\equiv n_{ct}^{\text{BK}}/n_{ct}^{\text{total}}$; and (iii) the *log ratio of breakthrough to imitative patents* $\equiv \ln(n_{ct}^{\text{BK}}+1) - \ln(n_{ct}^{\text{Mimic}}+1)$.

The first two outcomes are rates, not shares: because BK and Mimic are non-exhaustive top-decile labels, they do not sum to one, and each is centered near 0.10 by construction. A rate answers how frequently a city's patents reach either extreme, holding fixed the size of its portfolio; the third outcome answers the complementary question of which extreme dominates when a city does file at the tails. The two are distinct objects, and the composition-shift prediction requires both to move: a fall in the imitative rate paired with a rise in the breakthrough rate, and a rise in their log ratio. Reporting only one would leave open whether the portfolio shifted or merely thinned at both tails.

The first two outcomes use the first-difference (FD) specification (5), which eliminates all city-specific level effects through differencing, with $\gamma_t$ absorbing aggregate annual shocks. The third is estimated in levels with city and year fixed effects. Using total city patents $n_{ct}^{\text{total}}$ as the common denominator makes the two rate estimates directly comparable across cities of different size. Standard errors are clustered at the city level throughout.

Table 2: Industrial Policy and Innovation Composition: Main Results

| | (1) | (2) | (3) |
|---|---|---|---|
| | *Composition shift* | | *Innovation scale* |
| | $\Delta$ Mimic rate *(FD)* | $\Delta$ BK rate *(FD)* | ln(BK/Mimic) *(OLS)* |
| $\Delta \ln(\text{CumIP}_{ct}+1)$ | $-0.020^{*}$ | $0.033^{**}$ | |
| | (0.012) | (0.016) | |
| $\ln(\text{CumIP}_{ct}+1)$ | | | $0.317^{***}$ |
| | | | (0.051) |
| Observations | 2,056 | 2,056 | 3,458 |
| Within $R^2$ | 0.001 | 0.004 | 0.037 |
| City FE | *(by diff.)* | | ✓ |
| Year FE | ✓ | ✓ | ✓ |

*Notes:* Balanced city-year panel, 247 prefecture-level cities, 2000–2013. BK = breakthrough patent (top decile of the Kelly et al. (2021) innovation quality score within technology class and year, China-domestic threshold). Mimic = imitative patent (top decile of maximum cosine similarity to prior art, same scope). *Cols. (1)–(2):* first-difference (FD) specification $\Delta y_{ct} = \beta \Delta \ln(\text{CumIP}_{ct}+1) + \gamma_t + \varepsilon_{ct}$; city fixed effects absorbed by differencing; outcomes are shares of total city patents $n_{ct}^{\text{total}}$. *Col. (3):* OLS levels with city and year fixed effects; outcome is $\ln(n_{ct}^{\text{BK}}+1) - \ln(n_{ct}^{\text{Mimic}}+1)$. Standard errors are reported in parentheses. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, and 10% levels.

**Main results.** Table 2 reports the estimates. Columns (1)–(2) show that a one-log-unit increase in the cumulative IP stock lowers the imitative-patent rate by 2.0 percentage points (significant at the 10 percent level) and raises the breakthrough-patent rate by 3.3 percentage points (significant at the 5 percent level). The two rates move in opposite directions and by comparable magnitudes, which is what a reallocation across the tails implies and what a uniform contraction of extreme patenting would not. Column (3) confirms the shift from the complementary angle: IP accumulation raises the log ratio of breakthrough to imitative patents by 0.317, significant at the 1 percent level. Together the three estimates establish that IP reallocates the city-level patent portfolio from imitation toward breakthrough, the composition-shift prediction at the center of our theoretical framework.

**Dynamics and mechanism.** Figure 4 documents the aggregate time series: average log breakthrough-patent counts rise persistently across the sample period, while average log imitative-patent counts plateau after 2005, and the aggregate imitative-patent rate declines alongside the accelerating IP stock. Figure 5 decomposes the cross-sectional heterogeneity. Splitting cities at the median cumulative IP intensity, high-IP cities exhibit a persistently rising breakthrough-patent rate and falling imitative-patent rate relative to low-IP cities. The divergence emerges around 2003–2004, a two-to-three year lag behind policy accumulation. Panels C–D present distributed-lag impulse responses: the breakthrough-patent rate cumulates a response of roughly 2.7 percentage points over three years following a unit shock to $\Delta \ln(\text{CumIP}+1)$, and the imitative-patent rate mirrors this decline. The

response builds gradually rather than jumping, which points to persistent reallocation of inventor effort rather than a one-time change in filing behavior.

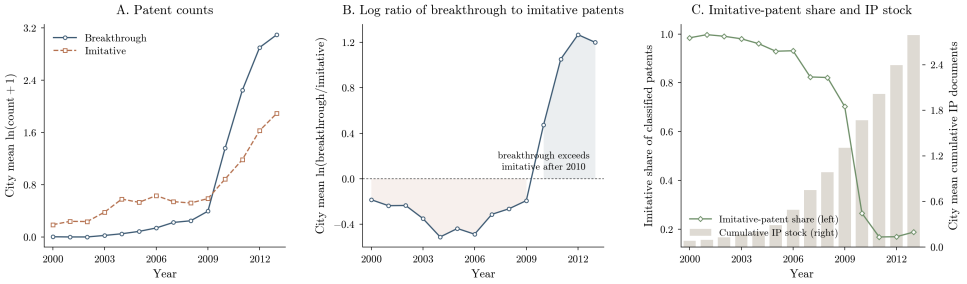



Figure 4: City-level patent composition and IP trends, 2000–2013.
*Notes:* Sample means across 247 prefecture-level cities. *Panel A:* mean log BK and log Mimic counts. *Panel B:* log BK/Mimic ratio; values above zero (post-≈2007) indicate BK output outpacing Mimic. *Panel C:* aggregate imitative-patent rate (left axis) and cumulative IP stock (right axis), both demeaned to 2000.

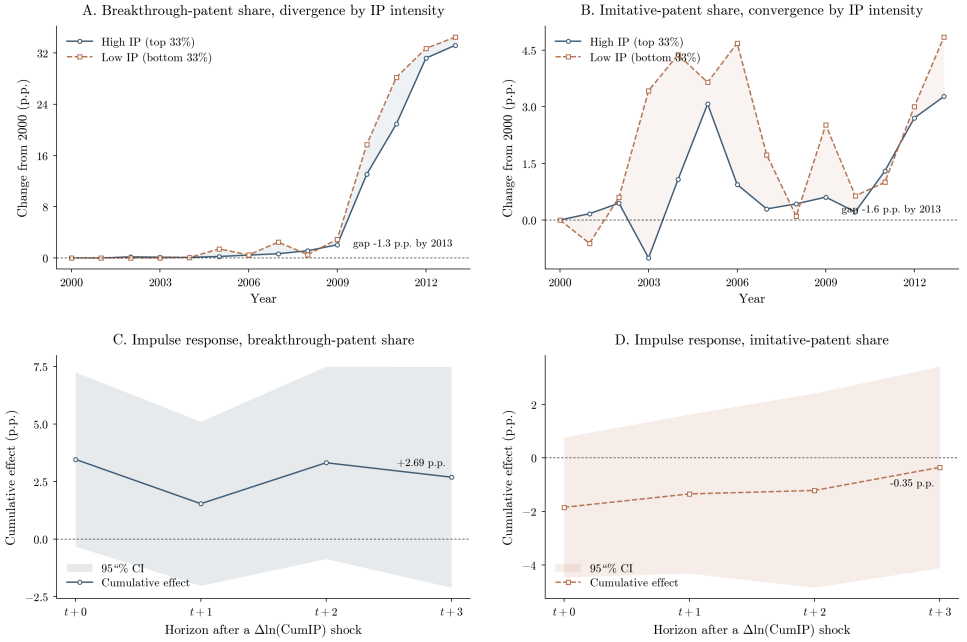



Figure 5: IP and the composition of innovation: dynamic evidence.
*Notes: Panels A–B:* change in breakthrough-patent rate (A) and imitative-patent rate (B) relative to 2000, for cities in the top and bottom tercile of mean ln(CumIP); shaded band marks the between-group gap. *Panels C–D:* cumulative impulse responses of Δbreakthrough-patent rate (C) and Δimitative-patent rate (D) to a unit shock in $\Delta \ln(\text{CumIP} + 1)$; bands = $\pm 1.96\,\widehat{\text{SE}}$ (delta method).

**Robustness.** Appendix Tables 15 and 16 verify that the three estimates in Table 2 are not sensitive to specification choices. Table 15 varies one dimension at a time: replacing year FE with province×year FE to absorb regional policy cycles (Panel A); controlling for total patent volume

$\ln n_{ct}^{\text{total}}$, so that the composition shift cannot be driven mechanically by overall patenting activity (Panel B); and winsorizing outcome and treatment variables at the 1st–99th percentiles (Panel C). All three estimates survive each check: the log ratio remains significant at the 1% level throughout, and the first-difference estimates preserve sign and magnitude.

Table 16 varies the treatment and the sample. Panel D replaces the cumulative IP stock with current-year IP documents; Panels E–F split the sample into early (2000–2006) and late (2007–2013) sub-periods; Panel G reassigns IP shocks randomly across cities within each year, 500 times. Under random assignment the estimated effects vanish—the permutation $p$-values place the true coefficients well outside the placebo distribution for both the breakthrough-patent rate and the log ratio—so the measured composition shift is not an artifact of how IP accumulation is distributed across the panel.

## 3.2 Micro Foundation: Imitation and Innovation Trade-off

The city-level patterns in Section 3.1 establish the macro regularity: IP accumulation shifts the patent portfolio from imitation toward breakthrough innovation. We now move from correlations to causal estimates of the *policy-by-knowledge* crowd-out channel at the firm level, the central prediction of our model. The analysis proceeds in three steps. We first establish the baseline pattern using our direct measure of local knowledge accessibility (Section 2.2.3). We then replicate and sharpen it using two independent variation strategies: high-speed rail (HSR) network expansion (Section 2.2.4) and import-embodied technology flows (Section 2.2.5). Finally, we address endogeneity through instrumental variables (Section 2.2.6).

One point about what this section does and does not establish. That subsidies raise innovation and reduce imitation on average is settled in Cai et al. (2025), on this same firm-patent-subsidy sample. We take that average response as given and ask a different question: how it varies with the knowledge a city can reach.

**Empirical specification.** Our workhorse specification compares subsidised and unsubsidised firms, before and after receipt, across cities that differ in knowledge access. We estimate:

$$Y_{ict} = \beta_1 D_{it} + \beta_2 K_{ct} + \beta_3 D_{it} \times K_{ct} + \Gamma X_{it} + \mu_i + \lambda_t + \varepsilon_{ict}, \tag{6}$$

where $Y_{ict} \in \{\text{TFP}, \text{TechInnov}, \text{Imitation}\}$ and $D_{it}$ equals one from the firm's first recorded subsidy receipt onward. $K_{ct}$ is accessible knowledge or the road-cost proxy, entered in deviation from its sample mean; $X_{it}$ includes firm age, firm history, export status, a relabeling-trend control, and city trends in population, per-capita GDP, and industrial structure; $\mu_i$ are firm fixed effects and $\lambda_t$ year fixed effects. Standard errors are clustered at the level at which treatment is assigned: the firm here and throughout the firm-level analysis, and the city in the regressions of Section 3.1.

The coefficient of interest is $\beta_3$. Since $K_{ct}$ enters in levels as well as in the interaction, it measures a difference in slopes: how much more a subsidy moves a firm's outcome when the city around it holds a deeper pool of accessible knowledge. Firms do not relocate in this panel, so $\mu_i$ absorbs every time-invariant city characteristic and a separate city effect would be collinear with it. Demeaning $K_{ct}$ leaves $\beta_1$ as the subsidy effect at average knowledge access.

### 2.2.3 General Knowledge Exposure

Table 3 reports estimates of equation (6). Three findings emerge.

Table 3: Innovation Subsidies, Knowledge Access, and Firm Innovation Outcomes

| | *Panel: Universe of Industrial Firms* | | | | | |
|---|---|---|---|---|---|---|
| | TFP | | Tech Innovation | | Imitation | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| If Subsidy | −0.054*** | −0.056*** | 0.049*** | 0.048*** | −0.072*** | −0.072*** |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.004) | (0.004) |
| If Subsidy × Accessible Knowledge | −0.069*** | | 0.048*** | | −0.068*** | |
| | (0.009) | | (0.005) | | (0.013) | |
| If Subsidy × Iceberg Cost | | 0.025*** | | −0.021*** | | 0.025*** |
| | | (0.004) | | (0.002) | | (0.005) |
| Observations | 2,122,753 | 2,122,753 | 2,699,985 | 2,699,985 | 2,605,266 | 2,605,266 |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Firm FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes:* The table reports estimates of equation (6). The sample is the universe of industrial firms in ASIEC. "If Subsidy" is an indicator equal to one from the firm's first recorded subsidy receipt onward. Odd columns interact the subsidy indicator with city-level accessible knowledge; even columns interact with the road-cost proxy labelled Iceberg Cost. Accessible Knowledge is constructed using the industry-adjusted TFP index described in Appendix D. Firm fixed effects absorb every time-invariant firm and city characteristic. Controls include firm age, firm history, export status, a relabeling-trend control, and city-level trends in log population, log per-capita GDP, and industrial output share. Standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% levels.

*Subsidies crowd out imitation and boost own innovation.* Across the full sample, a subsidy event is associated with lower TFP and lower imitation but higher own innovation. The subsidy indicator is negative and significant in the TFP and imitation columns (columns 1–2 and 5–6) and positive in the technology-innovation columns (columns 3–4). A subsidy event does not raise aggregate firm productivity on average; instead, it reallocates effort from imitation (which generates broad diffusion benefits) toward own R&D, consistent with the private-social wedge in our model.

*The crowd-out is amplified by accessible knowledge.* The subsidy-by-knowledge interaction is negative for TFP (column 1) and imitation (column 5) and positive for technology innovation (column 3), each significant at the 1 percent level. In knowledge-rich cities, subsidies amplify both the own-innovation gain and the imitation crowd-out, but the imitation channel dominates: the net effect on TFP is more negative where the accessible knowledge pool is deeper. Moving a subsidised firm from the 25th to the 75th percentile of the knowledge distribution amplifies the negative TFP impact by approximately 3 percentage points, a meaningful additional cost relative to the average subsidy effect.

*Transport costs produce the symmetric mirror image.* Columns (2), (4), and (6) replace accessible knowledge with the road-cost proxy. The interaction reverses sign throughout, positive and significant for TFP and imitation and negative for technology innovation. Where trade barriers are high and imitation opportunities are scarce, subsidies do least harm to aggregate productivity. This sign pattern follows from the research-allocation mechanism in Section 4.

Table 4: High-Speed Rail, Knowledge Access, and the Subsidy Crowd-Out

| | *Panel: Universe of Industrial Firms* | | | |
|---|---|---|---|---|
| | (1) Knowledge Access. | (2) TFP | (3) Tech Innovation | (4) Imitation |
| Railway Exposure Knowledge | 0.001*** | | | |
| | (0.000) | | | |
| If Subsidy | | −0.055*** | 0.043*** | −0.064*** |
| | | (0.003) | (0.001) | (0.004) |
| If Subsidy × Railway Exposure Knowledge | | −0.007*** | 0.012*** | −0.017*** |
| | | (0.001) | (0.001) | (0.001) |
| Observations | 2,711,489 | 2,123,347 | 2,700,739 | 2,606,002 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Firm FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |

*Notes:* The table reports estimates of equation (6) with Railway Exposure Knowledge as the spatial heterogeneity measure $K_{ct}$. Column (1) reports the first-order relationship between railway exposure and accessible knowledge. Columns (2)–(4) report the subsidy effect and its interaction with railway-based knowledge access on TFP, technology innovation, and imitation. Railway Exposure Knowledge is the interaction of a city's planned-network exposure with the log cumulative knowledge stock of hub cities connected to the HSR network by year $t$ (see Appendix E). Standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% levels.

### 2.2.4 High-Speed Rail and Knowledge Access

The results in Section 2.2.3 rely on our constructed measure of accessible knowledge, which may be subject to measurement error or reverse causality. We now exploit the staggered expansion of China's high-speed rail (HSR) network as a plausibly exogenous source of variation in knowledge access. The HSR instrument is described in detail in Appendix E; in brief, we interact a city's predetermined exposure to the planned HSR network (based on 1990s engineering plans) with the time-varying cumulative knowledge stock of hub cities that have opened HSR service. This produces a variable, Railway Exposure Knowledge, that captures the HSR-mediated expansion of a city's accessible knowledge pool.

Table 4 estimates equation (6) with Railway Exposure Knowledge as $K_{ct}$. Column (1) confirms the first link: a one-unit increase in Railway Exposure Knowledge raises city-level accessible knowledge significantly. Columns (2)–(4) recover the baseline pattern. The subsidy-by-railway-exposure interaction is negative and significant for TFP and imitation and positive and significant for technology innovation, the same sign pattern as Table 3, now traced to network expansion that follows corridors laid down decades earlier. As the HSR network reaches more knowledge-hub cities, the crowd-out cost of directing local firms toward own innovation rises with it.

### 2.2.5 Import-Embodied Technology Exposure

As a second independent robustness check, we replace railway-based knowledge access with city-level exposure to imports of high-technology products. Cities historically specialised in high-technology industries receive disproportionate knowledge inflows when global suppliers expand capacity for

Table 5: Import-Embodied Technology, Knowledge Access, and the Subsidy Crowd-Out

| | *Panel: Universe of Industrial Firms* | | | |
|---|---|---|---|---|
| | (1) Knowledge Access. | (2) TFP | (3) Tech Innovation | (4) Imitation |
| Import Exposure | 0.004*** | | | |
| | (0.000) | | | |
| If Subsidy | | −0.056*** | 0.045*** | −0.072*** |
| | | (0.003) | (0.002) | (0.005) |
| If Subsidy × Import Exposure | | −0.004*** | 0.010*** | −0.004** |
| | | (0.001) | (0.001) | (0.002) |
| Observations | 2,561,076 | 2,001,628 | 2,581,532 | 2,486,941 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Firm FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |

*Notes:* The table reports estimates of equation (6) with Import Exposure as the spatial heterogeneity measure $K_{ct}$. Column (1) reports the relationship between import exposure and accessible knowledge. Columns (2)–(4) report the subsidy effect and its interaction with import-embodied knowledge access on TFP, technology innovation, and imitation. Import Exposure is constructed from city-level imports of high-technology goods, weighted by baseline industrial composition (see Section 2.2.6). Standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% levels.

exogenous reasons. The formal Bartik construction is given in Section 2.2.6 and Appendix E.

Table 5 repeats the analysis using Import Exposure as $K_{ct}$. Column (1) confirms the first link: import exposure raises accessible knowledge, significant at the 1 percent level, consistent with embodied-technology imports transmitting foreign knowledge to Chinese firms. Columns (2)–(4) again recover the same qualitative pattern: subsidies reduce TFP and imitation while raising own innovation at conventional significance levels, and the interaction with import exposure amplifies the crowd-out of TFP and imitation while boosting own innovation. The consistency of results across three independent dimensions of knowledge access (direct TFP exposure, HSR connectivity, and import-embodied technology flows) suggests the mechanism is robust and not an artefact of any particular measurement choice.

### 2.2.6 Causal Identification

To address the potential endogeneity of city-level knowledge accessibility (through both international trade and high-speed rail), we construct two complementary instrumental variables. Both instruments combine a predetermined measure of a city's structural or geographic exposure with time-varying, externally driven shifts in knowledge availability, producing variation that is orthogonal to local economic conditions.

**Instrument I: Technology-Weighted Supply-Push Bartik IV.** Our first instrument follows the canonical supply-push Bartik design. Let $s_{c,k,t_0}$ denote the output share of industry $k$ in city $c$ in baseline year $t_0 = 2000$, and let $\varphi_{k,j,t_0} = M_{j\to\text{CHN},k,t_0} / \sum_{j'} M_{j'\to\text{CHN},k,t_0}$ denote China's baseline import-source share from country $j$ in product $k$. For each product $k$ and year $t$, the supply-push

shock is

$$g^{C}_{k,t} = \sum_{j} \varphi_{k,j,t_0} \times \Delta X_{k,j\to \mathrm{ADH8},t}, \tag{7}$$

where $\Delta X_{k,j\to \mathrm{ADH8},t}$ is supplier $j$'s change in exports to eight comparison countries, proxying for global supply capacity orthogonal to Chinese demand. Restricting to high-technology HS chapters (84, 85, 88, 90) that carry the most embodied knowledge and letting $\tau_k \in \{0,1\}$ indicate high-technology status, the technology-weighted shock is $g^{D}_{k,t} = \tau_k \times g^{C}_{k,t}$. The city-year instrument aggregates these shocks using predetermined industrial composition:

$$\mathrm{IV}^{\mathrm{Tech}}_{c,t} = \sum_{k\in\mathcal{K}} s_{c,k,t_0} \times g^{D}_{k,t}. \tag{8}$$

Cities historically specialised in high-technology industries experience larger knowledge inflows when global suppliers expand, making them more exposed to embodied foreign knowledge for reasons unrelated to local conditions.[3]

**Instrument II: Historical Railway × Hub Knowledge IV.** Our second instrument exploits two predetermined sources of variation. The first is the presence of functioning railways in 1933, constructed under the Republican-era national rail plan, as a proxy for structural predisposition toward major transport corridors. Let $\mathbf{1}[\mathrm{Railway}_{c,1933}]$ denote this time-invariant indicator. The second is the staggered opening of HSR among knowledge-hub cities. Let

$$\Psi_t = \ln\left(1 + \sum_{d\in\mathcal{H}} \widetilde{K}_{d,0} \cdot \mathbf{1}[\mathrm{HSR}_{dt} = 1]\right) \tag{9}$$

denote the cumulative knowledge stock accessible through HSR among hub cities active by year $t$. $\Psi_t$ is common to all cities in a given year and varies only with the national HSR rollout among hubs. Interacting the two sources:

$$\mathrm{IV}^{\mathrm{HSR}}_{c,t} = \mathbf{1}[\mathrm{Railway}_{c,1933}] \times \Psi_t. \tag{10}$$

Cities with pre-existing railways gain disproportionately larger hub-city knowledge access as the HSR network expands. Appendix E provides construction details and identification diagnostics. The first-stage $F$-statistics for both instruments consistently exceed 40.

**Two-Stage Least Squares Specification.** To instrument both knowledge access and its interaction with the subsidy indicator, we estimate the system:

$$\mathrm{Access}^{q}_{ct} = \pi_{11}\,\mathrm{IV}^{q}_{ct} + \pi_{12}\,P_{it}\,\mathrm{IV}^{q}_{ct} + \Gamma X_{it} + \mu_i + \lambda_t + u^{q}_{it}, \tag{11}$$

$$P_{it}\,\mathrm{Access}^{q}_{ct} = \pi_{21}\,\mathrm{IV}^{q}_{ct} + \pi_{22}\,P_{it}\,\mathrm{IV}^{q}_{ct} + \Gamma X_{it} + \mu_i + \lambda_t + v^{q}_{it}, \tag{12}$$

with second stage:

$$Y_{it} = \beta_1 P_{it} + \beta_2\,\widehat{\mathrm{Access}}^{q}_{ct} + \beta_3\,P_{it}\,\widehat{\mathrm{Access}}^{q}_{ct} + \Gamma X_{it} + \mu_i + \lambda_t + \varepsilon_{it}, \tag{13}$$

Table 6: IV Estimates: Import-Based Knowledge Access and Innovation Outcomes

| | First Stage | Second Stage | | |
|---|---|---|---|---|
| | (1) Import Exposure | (2) TFP | (3) Tech Innovation | (4) Imitation |
| If Subsidy | $-0.004^{*}$ | $0.051^{***}$ | $0.043^{***}$ | $0.066^{***}$ |
| | (0.002) | (0.008) | (0.005) | (0.010) |
| Bartik IV | $-0.0001^{***}$ | | | |
| | (0.000) | | | |
| If Subsidy × Bartik IV | $0.0001^{***}$ | | | |
| | (0.000) | | | |
| If Subsidy × Import Exposure $^{(IV)}$ | | $-0.189^{***}$ | $0.048^{***}$ | $-0.203^{***}$ |
| | | (0.011) | (0.007) | (0.014) |
| Import Exposure $^{(IV)}$ | | $-0.652^{***}$ | $-0.006$ | $-0.507^{***}$ |
| | | (0.090) | (0.017) | (0.062) |
| Observations | 1,966,886 | 1,966,886 | 2,554,967 | 2,429,444 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Firm FE | ✓ | ✓ | ✓ | ✓ |
| First-stage $F$-stat. | 163.3 | | | |
| Anderson–Rubin $p$-value | | $< 0.001$ | $< 0.001$ | $< 0.001$ |

*Notes:* The table reports 2SLS estimates of equations (11)–(13) using the technology-weighted supply-push Bartik IV (equation (8)). Column (1) is the first stage for Import Exposure. Columns (2)–(4) are second-stage estimates; $^{(IV)}$ denotes instrumented variables. The Anderson–Rubin Wald test $p$-value tests the null that all structural coefficients on the subsidised-firm interaction are zero. Standard errors are reported in parentheses. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, and 10% levels.

where $P_{it} = \mathbf{1}[\text{Subsidy}_{it}]$, $\mu_i$ are firm fixed effects, and $\lambda_t$ year fixed effects. Standard errors are clustered on the firm, as in equation (6). The coefficient $\beta_3$ is the IV estimate of the policy-by-knowledge crowd-out channel.

Table 6 reports the import-IV estimates. The first stage (column 1) is strong: the Bartik IV is significant and the $F$-statistic is 163.3, far above conventional weak-instrument thresholds. Comparing IV to OLS (Table 5), the interaction effects are substantially amplified. The IV estimates of the subsidy-by-import-exposure interaction are negative and significant at the 1 percent level for TFP and imitation, and positive and significant for technology innovation, all more than an order of magnitude larger than the corresponding OLS coefficients. The OLS attenuation is consistent with classical measurement error in import exposure and endogeneity of import composition toward locally demanded goods. The Anderson–Rubin Wald test rejects the null of zero structural interaction coefficients at $p < 0.001$ for all three outcome equations, providing robust inference under potential weak-instrument concerns.

Table 7 presents the railway-IV estimates. The first-stage $F$-statistic is 313.6. The second-stage interaction coefficients are negative and significant for TFP and imitation, and positive and significant for technology innovation, consistent with the import-IV results, confirming the same

[3]HS-6 shocks are mapped to two-digit CIC industries using a concordance; for many-to-one mappings, weights are split equally.

Table 7: IV Estimates: Railway-Based Knowledge Access and Innovation Outcomes

| | First Stage | Second Stage | | |
|---|---|---|---|---|
| | (1) Railway Exposure | (2) TFP | (3) Tech Innovation | (4) Imitation |
| If Subsidy | $-0.089^{***}$ | $-0.117^{***}$ | $0.059^{***}$ | $-0.057^{***}$ |
| | (0.007) | (0.006) | (0.002) | (0.005) |
| Railway IV | $0.063^{***}$ | | | |
| | (0.002) | | | |
| If Subsidy × Railway IV | $-0.029^{***}$ | | | |
| | (0.002) | | | |
| If Subsidy × Railway Exposure $^{(IV)}$ | | $-0.100^{***}$ | $0.051^{***}$ | $-0.034^{***}$ |
| | | (0.007) | (0.003) | (0.007) |
| Railway Exposure $^{(IV)}$ | | $-0.470^{***}$ | $0.064^{***}$ | $-0.015$ |
| | | (0.027) | (0.008) | (0.021) |
| Observations | 2,172,983 | 2,172,983 | 2,763,370 | 2,617,348 |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| Firm FE | ✓ | ✓ | ✓ | ✓ |
| First-stage $F$-stat. | 313.6 | | | |
| Anderson–Rubin $p$-value | | $< 0.001$ | $< 0.001$ | $< 0.001$ |

*Notes:* The table reports 2SLS estimates of equations (11)–(13) using the historical-railway × hub-knowledge IV (equation (10)). Column (1) is the first stage for Railway Exposure Knowledge. Columns (2)–(4) are second-stage estimates; $^{(IV)}$ denotes instrumented variables. The Railway IV is the interaction of an indicator for the presence of a functioning railway in city $c$ in 1933 with $\Psi_t$, the log cumulative knowledge stock of hub cities with active HSR service by year $t$. The Anderson–Rubin Wald test $p$-value tests the null that all structural interaction coefficients are zero. Standard errors are reported in parentheses. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, and 10% levels.

structural relationship between local knowledge access and the social cost of innovation subsidies.[4]

The OLS and IV estimates together establish three conclusions. *First*, innovation subsidies generate average-level reallocation from imitation toward own R&D that reduces aggregate TFP, consistent with a negative innovation-payoff wedge in the researcher-allocation margin. *Second*, the crowd-out is strongly amplified in knowledge-rich locations, where foregone imitation is most valuable. *Third*, two independent IV strategies confirm this interaction reflects a genuine structural relationship between the local knowledge environment and the social cost of innovation subsidies, not omitted city-level confounders or reverse causality.

# 4 Model

Section 3 documents one firm-level trade-off: a subsidy raises patenting and lowers the part of productivity that comes from outside learning, and the displacement is larger where accessible knowledge is richer. We embed this trade-off in a spatial, dynamic general-equilibrium economy in order to measure its aggregate and distributional consequences and to ask where, across space,

[4] Appendix H reports robustness checks including alternative first-stage specifications, alternative clustering levels, and placebo tests.

innovation subsidies belong. The model has three parts. A fixed local stock of researchers splits effort between creating new ideas and adopting ideas that diffuse in from other locations. A productivity frontier grows by accumulating both kinds of ideas. Eaton and Kortum (2002) trade links every location's frontier to every other location's market access and to the pool of ideas it can imitate. Industrial policy enters as a wedge in researchers' private return to innovation. Because the researcher pool is fixed, a subsidy cannot create research effort; it can only move effort from imitation to innovation. That single feature drives every result.

### 4.1 Environment

Time is discrete. The economy has $N$ locations indexed by $i$ and $n$; in the calibration these are thirty Chinese provinces plus an aggregate rest of the world. Each location houses two kinds of agents. Production workers are mobile and consume; researchers are immobile and own the local technology. Each variety is produced with a Fréchet-distributed productivity draw, the standard primitive that delivers the Eaton and Kortum (2002) trade structure and an idea frontier that accumulates additively. Households consume a CES bundle of varieties with elasticity $\sigma > 1$,

$$\mathcal{C}_{i,t} = \Big( \int_{\omega} c_{i,t}(\omega)^{\frac{\sigma-1}{\sigma}} \, d\omega \Big)^{\frac{\sigma}{\sigma-1}}. \tag{14}$$

The aggregate productivity frontier of location $n$, denoted $A_{n,t}$, is the scale of its idea distribution. Two activities raise it: innovation creates new domestic ideas, and imitation adopts foreign ideas that have diffused into $n$.

### 4.2 Workers and Migration

Production workers supply one unit of labor inelastically to local firms. Given wage $w_{i,t}$, price index $P_{i,t}$, and a residential amenity $B_{i,t}$, flow utility is $\log(w_{i,t}/P_{i,t}) + B_{i,t}$. At the end of each period a worker in $i$ chooses next period's location, paying a bilateral cost $m_{in,t}$ and drawing i.i.d. Type-I extreme-value location tastes. With discount factor $\beta \in (0,1)$, the value before the location choice is

$$V_{i,t} = \log\Big(\tfrac{w_{i,t}}{P_{i,t}}\Big) + B_{i,t} + \beta \, \mathbb{E}_{\varepsilon}\Big[ \max_{n} \{ V_{n,t+1} - m_{in,t} + \varepsilon_{in,t} \} \Big]. \tag{15}$$

The taste shocks have scale $\zeta > 0$, so the share of workers who move from $i$ to $n$ is

$$\mu_{in,t} = \frac{\exp\big((V_{n,t+1} - m_{in,t})/\zeta\big)}{\sum_{h} \exp\big((V_{h,t+1} - m_{ih,t})/\zeta\big)}. \tag{16}$$

Population grows at gross rate $g > 1$ through local births, so the distribution of production workers evolves as $L^{P}_{n,t+1} = \sum_{i} \mu_{in,t} \, g \, L^{P}_{i,t}$. The amenity $B_{i,t}$ shifts residence value only; it does not touch production, research, trade, or diffusion, and we use it solely to let the calibrated path reproduce the observed population distribution once migration costs are disciplined by migration flows.

### 4.3 Researchers, Innovation, and Imitation

Each location is endowed with an exogenous, immobile measure $\bar{L}^{R}_{n,t}$ of researchers. This is the model's central friction: the research stock is fixed within the period, so the only margin a subsidy can move is its split between innovation and imitation,

$$L^{\mathrm{I}}_{n,t} + L^{\mathrm{M}}_{n,t} = \bar{L}^{R}_{n,t}, \qquad L^{\mathrm{I}}_{n,t}, L^{\mathrm{M}}_{n,t} \geq 0. \tag{17}$$

Innovation effort produces new ideas with intensity

$$R_{n,t} = a_n^{\mathrm{I}} (L_{n,t}^{\mathrm{I}})^{\nu}, \qquad \nu \in (0,1), \tag{18}$$

where $a_n^{\mathrm{I}}$ is innovation productivity and $\nu$ is the curvature of research effort. Imitation operates on the *imitable pool*: foreign ideas that have arrived in $n$ before $t$ through trade but are not yet known locally. The pool is built from the past innovation of *other* locations, weighted by how far their ideas have diffused along the trade network,

$$A_{n,t}^{\mathrm{pool}} = \sum_{i \neq n} \sum_{s<t} (1 - \sigma_{in}^{s,t-1})\, G_{in}^{s,t-1}\, R_{i,s}, \tag{19}$$

where $\sigma$ and $G$ are diffusion objects determined by bilateral trade exposure; Appendix F.1 derives both objects and the pool equation. The sum runs over $i \neq n$: a location's own innovation seeds *other* provinces' imitable pools, never its own. This cross-location externality is the second force in the analysis. Imitation effort $a_n^{\mathrm{M}}(L_{n,t}^{\mathrm{M}})^{\nu}$ converts pool ideas into adopted ideas with learning elasticity $\eta \in (0,1)$, delivering an imitation frontier

$$A_{n,t}^{\mathrm{im}} = \Gamma(1 - \tfrac{\theta}{\vartheta})\, d\, (a_n^{\mathrm{M}})^{\eta} (A_{n,t}^{\mathrm{pool}})^{1-\eta} (L_{n,t}^{\mathrm{M}})^{\nu\eta}, \tag{20}$$

with $d$ a learning-efficiency scalar and $\Gamma(1 - \theta/\vartheta)$ a Fréchet constant. Imitation is decreasing in own effort curvature and increasing in the local pool: where more foreign knowledge has arrived, the same research time yields more adoption.

### 4.4 The Productivity Frontier and Trade

Because the frontier is the scale of a Fréchet idea distribution, combining idea sources *adds* their scales: the maximum of independent Fréchet draws is Fréchet with scale equal to the sum. The frontier therefore evolves *additively* in the two idea flows,

$$A_{n,t} = A_{n,t-1} + \Gamma(1 - \tfrac{\theta}{\vartheta}) R_{n,t} + A_{n,t}^{\mathrm{im}}. \tag{21}$$

Appendix F.2 derives this recursion exactly from the marked-Poisson idea process and the product of the three Fréchet cumulative distribution functions. This additive law is not a modeling convenience: it is forced by the same Fréchet structure that delivers the trade shares below. It also disciplines how we read the data. The firm productivity decomposition in Section 2 is in logs, whereas the frontier in (21) is additive in Fréchet scales. We therefore use the estimated policy slopes to identify responses and use the exponentiated firm components only to construct relative spatial level moments after aggregation. A common scalar, identified from the TFP policy-by-access sign change, maps those empirical moments into the model's additive units; the raw log residual is never treated as an additive frontier increment. Section 5 describes this mapping.

Trade follows Eaton and Kortum (2002) with correlated productivity across sources (Lind and Ramondo, 2023). With iceberg cost $\tau_{in,t}$, wage $w_{n,t}$, and trade elasticity $\theta$, define $x_{in,t} \equiv A_{n,t}(\tau_{in,t} w_{n,t})^{-\theta}$. Importer $i$'s expenditure share on goods from $n$ is

$$\lambda_{in,t} = \frac{x_{in,t}\, \mathcal{G}_n(x_{i,t})}{\mathcal{G}(x_{i,t})}, \tag{22}$$

where $\mathcal{G}$ is the correlation generator and $\mathcal{G}_n$ its derivative in coordinate $n$ (the no-correlation case gives the textbook share $x_{in} / \sum_h x_{ih}$). Trade thus links every location's frontier to world demand, and through (19) it also governs which foreign ideas each location can imitate.

### 4.5 Industrial Policy as a Payoff Wedge

Researchers allocate effort to equate the private marginal value of innovation and imitation. Industrial policy distorts this choice. We let an innovation-oriented policy raise the private payoff of innovation by a factor $\Omega_{n,t} \geq 0$, leaving the true technologies (18)–(20) unchanged. The interior allocation then satisfies

$$\Omega_{n,t}\, a_n^{\mathrm{I}} (L_{n,t}^{\mathrm{I}})^{\nu-1} = d\,\eta\, (a_n^{\mathrm{M}})^{\eta} (A_{n,t}^{\mathrm{pool}})^{1-\eta} (L_{n,t}^{\mathrm{M}})^{\nu\eta-1}, \tag{23}$$

together with the stock constraint (17). The policy maps to $\Omega$ through the observed subsidy exposure $Z_{n,t}$,

$$\Omega_{n,t} = \bar{\Omega}_n \exp(\chi_n Z_{n,t}), \qquad \chi_n > 0, \tag{24}$$

where $\bar{\Omega}_n$ is the baseline distortion and $\chi_n$ the province-specific policy pass-through. The wedge enters (23) but not output: $\Omega$ multiplies the private *incentive* to innovate, not the productivity of innovation effort. An equivalent reading is a subsidy that lowers the relative private cost of innovation by the factor $1/\Omega$; under the fixed research stock the two are observationally identical. The crucial implication is that a higher $\Omega$ cannot raise total research output. It can only shift researchers from imitation toward innovation. Two forces follow directly. First, an *innovation-reallocation* effect: $L^{\mathrm{I}}$ rises and $L^{\mathrm{M}}$ falls, so own imitation drops, and the loss is large where the imitable pool is rich because that is where the marginal researcher in imitation was most productive. Second, a *knowledge-pool misallocation* effect: because the imitable pool varies across space, the same wedge destroys more imitation precisely where imitation is the more valuable margin. Section 6 quantifies both forces.

### 4.6 Equilibrium and Balanced Growth

Production profits accrue to the immobile researchers, the idea owners, so per-researcher income equals local profit per researcher. Goods-market clearing under monopolistic competition gives $\theta\,\Pi_{n,t}^{P} = w_{n,t} L_{n,t}^{P}$ and the standard revenue-equals-expenditure condition; labor-market clearing requires $L_{n,t}^{R} + L_{n,t}^{P} = L_{n,t}$. A sequential equilibrium is a path of wages such that, in every period, trade shares (22), the price index, the research allocation (23), goods- and labor-market clearing, the migration block (15)–(16), and the frontier law (21) all hold. On a balanced growth path all real quantities grow at the common rate $g$; we conduct the quantitative analysis on the detrended stationary equilibrium and report the hat-algebra in the Appendix.

## 5 Calibration

We calibrate a five-year stationary equilibrium with thirty Chinese provinces and a rest-of-world region. The model's exogenous parameters fall into six groups: trade and research-technology parameters, intertemporal and spatial-mobility parameters, policy and learning parameters, research fundamentals, location endowments and accounting fundamentals, and trade and migration fundamentals. Table 8 summarizes the identification of every parameter. The underlying moments, identifying equations, and numerical inversions are deferred to Appendix G.

**Trade and research-technology parameters: $\{\theta, \vartheta, \nu, d\}$.** The Fréchet shape parameter $\theta$ governs comparative advantage and equals the trade elasticity in the independent Eaton and Kortum (2002) limit. We set $\theta = 2.65$ using the aggregate trade elasticity in Lind and Ramondo

(2023). The applicability parameter $\vartheta$ governs how strongly the productivity of a common idea co-moves across locations. We set $\vartheta = 5.00$ using their within-factor elasticities for innovation-intensive manufactured goods, which implies a one-nest correlation of $1-\theta/\vartheta = 0.47$. The research curvature $\nu$ determines diminishing returns to research labor in both innovation and imitation. We set $\nu = 0.76$ following Xiang (2025) and Bloom et al. (2020). The common learning efficiency $d$ scales the probability of acquiring an externally available idea. Because only its composite with location-specific imitation productivity is separately identified, we normalize $d = 1$ and recover imitation productivity conditional on this normalization.

**Intertemporal and spatial-mobility parameters: $\{\beta, g, \zeta\}$.** The discount factor $\beta$ governs the valuation of future researcher income and enters the worker and researcher dynamic problems. We use an annual value of 0.95, which gives $\beta = 0.95^5 = 0.774$ over the model's five-year period. The common growth factor $g$ governs balanced-growth population and researcher growth in every location. Working-age population growth between the 2000 and 2010 censuses gives $g = 1.066$ over five years. The migration taste-shock scale $\zeta$ determines the responsiveness of migration to destination value. Following Tombe and Zhu (2019), we map their migration elasticity of 1.5 into $\zeta = 1/1.5 = 0.667$. Using a common $g$ and $\zeta$ ensures that the demographic and migration blocks are measured over the same five-year horizon.

**Policy and learning parameters: $\{\eta, \chi_n, \bar{\Omega}_n\}$.** The learning elasticity $\eta$ determines the weight placed on local imitation capacity relative to the externally available idea stock in the imitation technology. It is identified by the response of imitation to an exogenous change in accessible external knowledge, giving $\eta = 0.909$. The policy pass-through $\chi_n$ is the semi-elasticity of the private innovation-payoff wedge with respect to the subsidy shock. It is identified from the innovation and imitation responses to that shock and ranges from 0.0257 to 0.0662 across Chinese provinces. The baseline wedge $\bar{\Omega}_n$ measures the private return to innovation relative to imitation before the counterfactual policy change. We recover it from the research-allocation condition after expressing the two research activities in a common productivity unit. The resulting wedge ranges from 0.246 to 6.588. Appendix G reports the coefficient-to-parameter mapping and the common-unit inversion.

**Research fundamentals: $\{\bar{a}_n^I, a_n^M\}$.** Innovation productivity $\bar{a}_n^I$ measures the amount of new frontier knowledge generated by a given innovation-labor input. We recover it separately for each location by inverting the innovation technology at the calibrated research allocation. Imitation productivity $a_n^M$ measures a location's capacity to absorb ideas that have arrived but are not yet locally known. We recover it from the imitation technology after solving the diffusion fixed point for the stock of externally available ideas. These inversions ensure that the calibrated research technologies and the research-allocation condition hold exactly at the reference equilibrium. Both productivities are treated as time-invariant fundamentals and remain fixed in every policy and trade-cost counterfactual. Appendix G gives the full inversion and the separate treatment of the rest-of-world region.

**Location endowments and accounting fundamentals: $\{\bar{L}_n^R, \bar{L}_n^P, D_n\}$.** The researcher endowment $\bar{L}_n^R$ fixes the local research-labor stock that is allocated between innovation and imitation. For Chinese provinces, it is measured from science-and-technology personnel in the 2010 city statistical yearbooks and expressed in the same labor unit as production employment. The reference production-labor endowment $\bar{L}_n^P$ is constructed from total employment in the same yearbooks and normalized so that Chinese production labor sums to one. Rest-of-world researcher and production-labor endowments use World Bank researcher and labor-force totals under the same normalization. The net trade deficit $D_n$ is constructed as expenditure minus revenue in the com-

bined interregional and international input-output accounts. It allows the reference equilibrium to reproduce the observed trade imbalance rather than imposing balanced trade province by province.

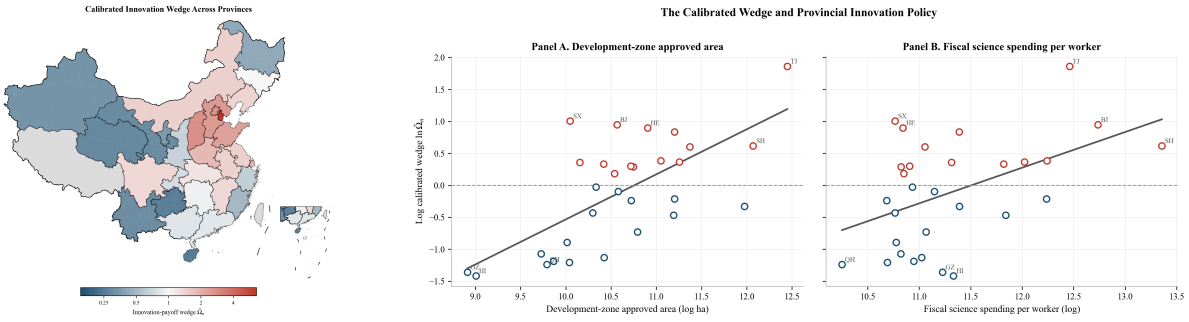
**Trade and migration fundamentals: $\{\tau_{on}, m_{in}, B_n\}$.** Directed iceberg trade costs $\tau_{on}$ govern the cost of shipping from origin $o$ to destination $n$. Conditional on the recovered technology fundamentals, we choose the full directed matrix to reproduce bilateral expenditure shares under the correlated-GEV trade system. Symmetric migration costs $m_{in}$ measure the utility cost of moving between Chinese provinces and are recovered from reciprocal bilateral migration odds in the 2010 census, given $\zeta$. Residential amenities $B_n$ capture persistent location utility not explained by real income or migration costs. We choose them so that the stationary migration equilibrium reproduces the observed production-labor distribution while satisfying the worker Bellman equation. The resulting $\tau_{on}$, $m_{in}$, and $B_n$ are held fixed unless a counterfactual directly changes trade costs. Appendix G reports all three inversions and their numerical implementation.

Table 8: Parameters and Identification

| Parameter | Role | Identification | Estimate |
|---|---|---|---|
| *Trade and research-technology parameters.* | | | |
| $\theta$ | Trade elasticity | Lind and Ramondo (2023) | 2.65 |
| $\vartheta$ | Idea applicability | Lind and Ramondo (2023) | 5.00 |
| $\nu$ | Research curvature | Xiang (2025); Bloom et al. (2020) | 0.76 |
| $d$ | Common learning efficiency | Normalization | 1 |
| *Intertemporal and spatial-mobility parameters.* | | | |
| $\beta$ | Five-year discount factor | Annual value 0.95 compounded over five years | 0.774 |
| $g$ | Five-year growth factor | 2000 and 2010 population censuses | 1.066 |
| $\zeta$ | Migration taste-shock scale | Tombe and Zhu (2019) | 0.667 |
| *Policy and learning parameters.* | | | |
| $\eta$ | Imitation learning elasticity | Knowledge-access response of imitation | 0.909 |
| $\chi_n$ | Policy pass-through to the innovation wedge | Policy responses of innovation and imitation | 0.0257–0.0662 |
| $\bar{\Omega}_n$ | Baseline innovation-payoff wedge | Research-allocation condition | 0.246–6.588 |
| *Research fundamentals.* | | | |
| $\bar{a}_n^I$ | Innovation productivity | Innovation-technology inversion | Location specific |
| $a_n^M$ | Imitation productivity | Imitation-technology inversion | Location specific |
| *Location endowments and accounting fundamentals.* | | | |
| $\bar{L}_n^R$ | Researcher endowment | 2010 research personnel | Location specific |
| $\bar{L}_n^P$ | Reference production-labor endowment | 2010 employment | Location specific |
| $D_n$ | Net trade deficit | Input-output expenditure less revenue | Location specific |
| *Trade and migration fundamentals.* | | | |
| $\tau_{on}$ | Directed iceberg trade cost | Correlated-GEV trade-share inversion | Bilateral |
| $m_{in}$ | Symmetric migration cost | Reciprocal 2010 census migration odds | Bilateral |
| $B_n$ | Residential amenity | Stationary production-labor distribution | Location specific |

**Calibrated wedge and external validation.** The calibrated wedge lies below one in sixteen provinces and above one in fourteen. A value above one means that private incentives reward innovation too strongly relative to imitation at the margin; it does not simply indicate that a province receives more subsidy spending. Figure 6 shows high wedges concentrated in North China and lower wedges across much of the West and Southwest. Provinces with larger development-zone programs and greater public science spending per worker also tend to have higher innovation-payoff wedges. These policy measures are excluded from the calibration, so the comparison is an external check rather than an identifying moment.

The inversion reproduces the targeted trade shares, wages, production-labor distribution, and innovation and imitation flows. These exact matches establish internal consistency because the corresponding fundamentals are recovered from those moments. The reduced-form knowledge and policy gradients and the independent policy comparison in Figure 6 provide the non-targeted checks on the mechanism.




A. Geographic distribution

B. External policy validation

Figure 6: The Calibrated Innovation Wedge

*Notes:* Panel A maps the calibrated innovation wedge $\bar{\Omega}_n$ on a log scale centered at $\bar{\Omega}_n = 1$. Red provinces are over-incentivized toward innovation and blue provinces are under-incentivized. Panel B compares the log wedge with development-zone approved area and fiscal science-and-technology spending per worker. These policy measures are excluded from the calibration. Grey areas are outside the thirty-province sample.

# 6 Counterfactuals

We use the calibrated economy to quantify how innovation policy and market access alter the allocation of research across space. The first two counterfactuals raise one province's innovation wedge by 10 percent or reduce its bilateral trade costs by 10 percent, one province at a time. The third counterfactual chooses the full province-specific wedge schedule. We re-solve the stationary general equilibrium in every experiment and report production-labor-weighted aggregate productivity, $\Delta \ln \mathrm{TFP} = \sum_n s_n^P \Delta \ln A_n$, together with consumption-equivalent worker welfare.

## 6.1 The Crowd-Out Mechanism

We first raise each province's innovation wedge by 10 percent, one province at a time. Figure 7 shows the central trade-off. Innovation rises in every province by 4.5 to 21.1 percent, while imitation falls by 8.1 to 17.0 percent. These opposing responses leave the local productivity frontier higher in fifteen provinces and lower in fifteen. A common innovation incentive therefore has no common productivity effect across space.

The sign of the productivity response follows the initial research distortion. Because the local researcher stock is fixed, a higher wedge moves researchers from imitation to innovation and bounds the imitation response by $-\nu\eta/(1-\nu\eta)$. The innovation gain is largest where the initial innovation base is small, whereas the opportunity cost of displaced imitation is largest where the imitable knowledge pool is rich. The North China corridor has high calibrated wedges, so the marginal product of additional innovation is low relative to the foregone return from imitation. Tianjin illustrates this case, with a frontier decline of 8.8 percent. In much of western China the wedge is below one, and the same policy corrects the research allocation. Hainan records a frontier gain of 8.8 percent. The marginal-effect map is consequently close to the inverse of the calibrated-wedge map in Figure 6A.

Frontier distance summarizes part, but not all, of this heterogeneity. Figure 8 shows that provinces farther from the world frontier tend to gain more from an innovation incentive, with a correlation of 0.46. The imitation response is only weakly related to frontier distance because it also depends on the initial wedge and the local knowledge pool. The policy also affects other provinces

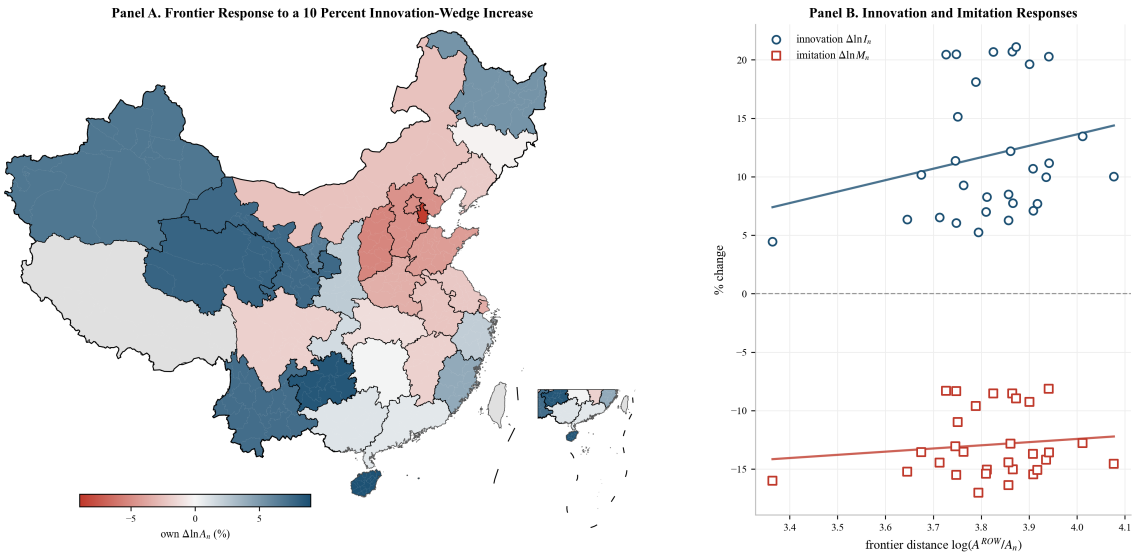



Figure 7: A Marginal Increase in the Innovation Wedge

*Notes:* Each province's innovation wedge $\Omega_n$ is raised by 10 percent, one province at a time, with the stationary general equilibrium re-solved after each shock. The map colors each province by the response of its own frontier $\Delta \ln A_n$. Blue denotes a frontier increase, red denotes a frontier decline, and grey denotes areas outside the calibration sample. The right panel reports innovation and imitation responses against model frontier distance $\log(A^{ROW}/A_n)$. Innovation $I_n$ rises everywhere and imitation $M_n$ falls everywhere.

through diffusion. Figure 11 shows that the average frontier response outside the treated province is positive in twenty-three of the thirty experiments, including cases in which the treated province itself loses. The distinction between the local reallocation cost and the external diffusion benefit is central to the optimal policy below.

The China-wide welfare response to the same 10 percent wedge increase also sorts sharply with the province's initial research distortion. Panel A of Figure 13 plots this response against the calibrated innovation wedge. All sixteen provinces with $\bar{\Omega}_n < 1$ generate welfare gains; all fourteen with $\bar{\Omega}_n > 1$ generate losses. The gains range from 0.0044 percent of consumption-equivalent welfare in Jilin to 0.078 percent in Zhejiang, and the largest loss is 0.114 percent in Shandong. Jilin is the only province for which China-wide welfare and the province's own frontier move in opposite directions. Its calibrated wedge is 0.994. The experiment lowers its own frontier by 0.217 percent and China-wide TFP by 0.0020 percent, but raises China-wide worker welfare by 0.0044 percent. Positive welfare spillovers to other provinces account for most of this gain. A wedge of one implements the locally efficient research allocation; the welfare response also reflects diffusion gains elsewhere and the wedges prevailing in other provinces.

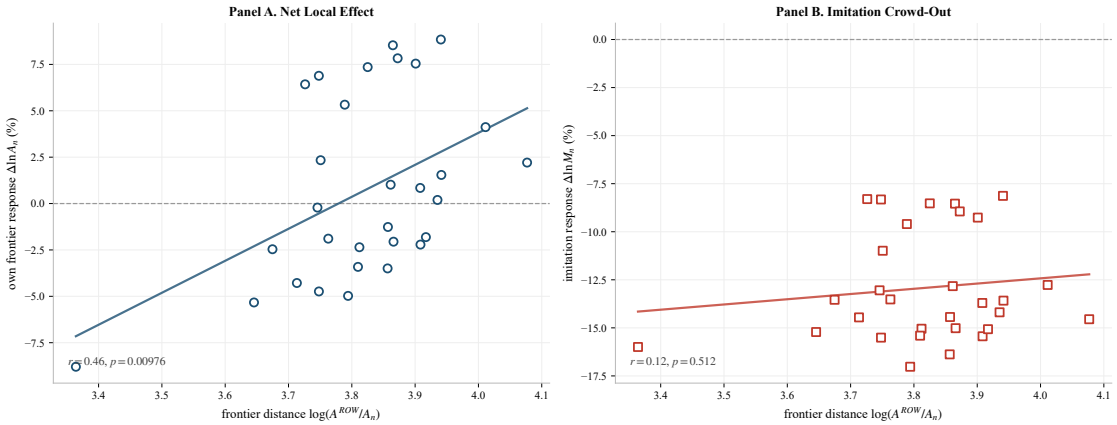



Figure 8: Heterogeneity in the Marginal Innovation-Wedge Effect

*Notes:* Each point represents one province-specific counterfactual from Figure 7. The horizontal axis reports model frontier distance $\log(A^{ROW}/A_n)$. Panel A reports the province's frontier response and Panel B its imitation response. Black lines are OLS fits.

## 6.2 The Market-Access Margin

We next reduce each province's bilateral trade costs by 10 percent in both directions, one province at a time. Figure 9 shows that the channel responses reverse. Imitation rises in every province by 1.0 to 5.6 percent, while innovation falls by 0.3 to 4.3 percent. The local productivity frontier rises in twenty-one of the thirty provinces. A simultaneous 10 percent reduction in bilateral trade costs raises aggregate TFP by 1.03 percent and worker welfare by 4.04 percent.

The mechanism operates through the diffusion network. Lower trade costs increase exposure to foreign ideas and expand the imitable pool in (19). The richer pool raises the return to imitation, which reallocates researchers away from innovation. Unlike the policy wedge, market access does not merely change the allocation of a fixed research stock. It also changes the stock of ideas available for adoption. This expansion in external knowledge explains why most local frontiers rise even though innovation falls everywhere.

The spatial incidence is correspondingly reversed. Across provinces, the correlation between the frontier responses to the policy and trade-cost experiments is $-0.90$. Shandong, Tianjin, and Hebei gain the most from improved market access and lose from a further innovation incentive. In these provinces, greater access to imitation inputs offsets an initially high innovation wedge. The nine provinces that lose from lower trade costs begin with an under-incentivized innovation margin, so the loss from reallocating researchers away from innovation exceeds the benefit of the richer pool. Figure 10 confirms that this response is not summarized by frontier distance alone. Network position determines which ideas arrive and therefore remains an independent source of heterogeneity.

Market access also generates spillovers beyond the treated province. Figure 11 compares these effects with those from the innovation-wedge experiment. The external frontier response is positive in eighteen of the thirty trade-cost experiments but smaller than the local response. Its weak relationship with frontier distance reinforces the network interpretation. A change in one province's trade costs alters the matrix through which ideas reach all other provinces, not only the treated province's distance from the world frontier.

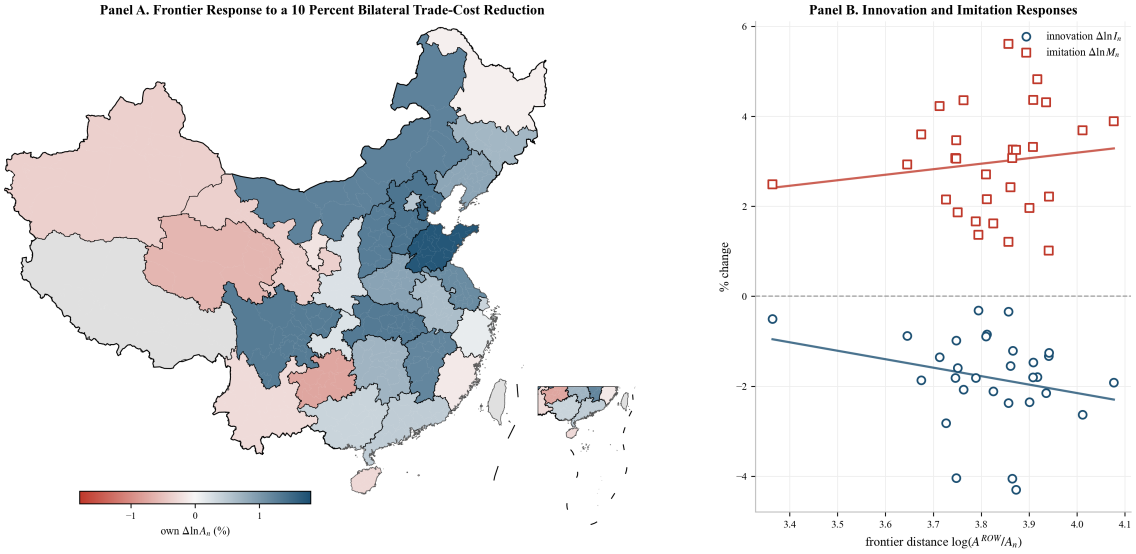



Figure 9: A Marginal Reduction in Bilateral Trade Costs

*Notes:* Each province's bilateral trade costs $\tau_{in}$ are cut by 10 percent in both directions, one province at a time, with the stationary general equilibrium re-solved after each shock. The map colors each province by the response of its own frontier $\Delta \ln A_n$. Blue denotes a frontier increase, red denotes a frontier decline, and grey denotes areas outside the calibration sample. The right panel reports imitation and innovation responses against model frontier distance $\log(A^{ROW}/A_n)$. Imitation $M_n$ rises everywhere and innovation $I_n$ falls everywhere, reversing the channel pattern in Figure 7.

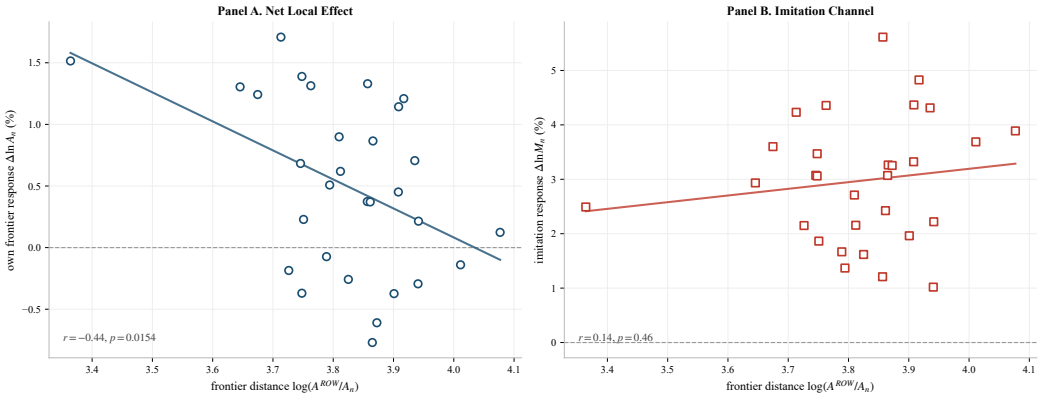
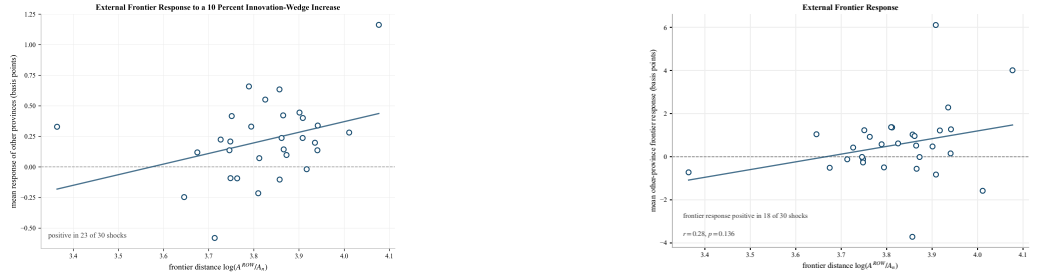



Figure 10: Heterogeneity in the Marginal Market-Access Effect

*Notes:* Each point represents one province-specific counterfactual from Figure 9. The horizontal axis reports model frontier distance $\log(A^{ROW}/A_n)$. Panel A reports the province's frontier response and Panel B its imitation response. Black lines are OLS fits.




A. Innovation-wedge increase

B. Trade-cost reduction

Figure 11: External Frontier Responses to Local Shocks

*Notes:* Each point is a province-specific counterfactual. Panel A raises that province's innovation wedge by 10 percent. Panel B cuts all bilateral trade costs into and out of that province by 10 percent. The horizontal axis is province $n$'s model frontier distance, $\log(A^{ROW}/A_n)$. The vertical axis reports the unweighted average frontier response of the other twenty-nine provinces, $\frac{1}{29}\sum_{m \neq n} \Delta \ln A_m$, in basis points.

### 6.3 The Optimal Subsidy and the Cost of Misallocation

We finally choose the full province-specific wedge schedule. The planner takes all calibrated fundamentals as given and internalizes the effects of each province's innovation on knowledge diffusion elsewhere. In the baseline exercise the wedge changes private research incentives without using real resources, so the optimum isolates the spatial allocation of innovation incentives from their fiscal financing. We solve separately for the schedules that maximize worker welfare and aggregate TFP. Appendix G describes the numerical optimization.

The optimal policy is mild, spatially differentiated, and far less dispersed than the calibrated policy. Under the welfare objective, optimal subsidy rates range from 5 to 33 percent and the research-weighted mean wedge equals 1.18. Replacing the calibrated schedule with this allocation raises aggregate manufacturing TFP by 11.26 percent and worker welfare by 4.25 percent. Under the TFP objective, subsidy rates range from −3 to 36 percent, aggregate TFP rises by 11.52 percent, and welfare rises by 4.09 percent. Table 9 shows that the two objectives deliver similar allocations and aggregate gains. The excessive spatial dispersion of the calibrated policy accounts for most of the loss.

Table 9: The Optimal Wedge under Two Planner Objectives

| Planner objective | Subsidy rate range | $\Delta \ln \text{TFP}$ | $\Delta W$ |
|---|---|---|---|
| Maximize welfare | +5% to +33% | +11.26% | +4.25% |
| Maximize TFP | −3% to +36% | +11.52% | +4.09% |

Notes: Each schedule chooses the province-specific wedge $\{\bar{\Omega}_n\}$ to maximize the stated general-equilibrium objective. The subsidy rate is $\bar{\Omega}_n - 1$. The calibrated wedge ranges from 0.25 to 6.59. Changes in TFP are production-labor weighted, and $\Delta W$ is the consumption-equivalent change in worker welfare. All changes are relative to the calibrated economy.

Welfare-optimal subsidy rates tend to decline with distance from the world frontier; their correlation with frontier distance is −0.40. Diffusion benefits also contribute to the cross-province variation in optimal rates.

The optimal schedule balances a local reallocation cost against an external diffusion benefit. Figure 12 measures both margins from the undistorted allocation, $\bar{\Omega}_n = 1$. At this point the research allocation maximizes the province's own frontier, so a marginal innovation wedge weakly lowers local productivity. The magnitude of this loss is smaller near the world frontier. The external benefit arises because additional innovation expands other provinces' imitable pools. This external benefit has a correlation of 0.72 with trade-weighted diffusion reach and 0.07 with frontier distance. Both margins shape the schedule. The local loss varies with frontier distance, and the external benefit is closely related to diffusion reach. The optimal schedule therefore targets provinces that combine a small local reallocation loss with broad diffusion reach. Jiangsu and Shanghai are leading examples.

The calibrated wedge ranges from 0.25 to 6.59; its log dispersion is sixteen times the welfare-optimal value, which ranges from 1.05 to 1.33. Panel B of Figure 13 displays the welfare-optimal rate and the gap between the calibrated and optimal rates. The largest positive gaps are concentrated in the North China corridor, while the largest negative gaps lie in western and southwestern provinces. Tianjin has a 25.0 percent welfare-optimal rate and a 558.8 percent calibrated implicit rate, a gap of 533.8 percentage points. Tianjin's exceptionally high calibrated wedge accounts for this extreme gap. Hainan lies 94 percentage points below its optimum. Additional innovation in-

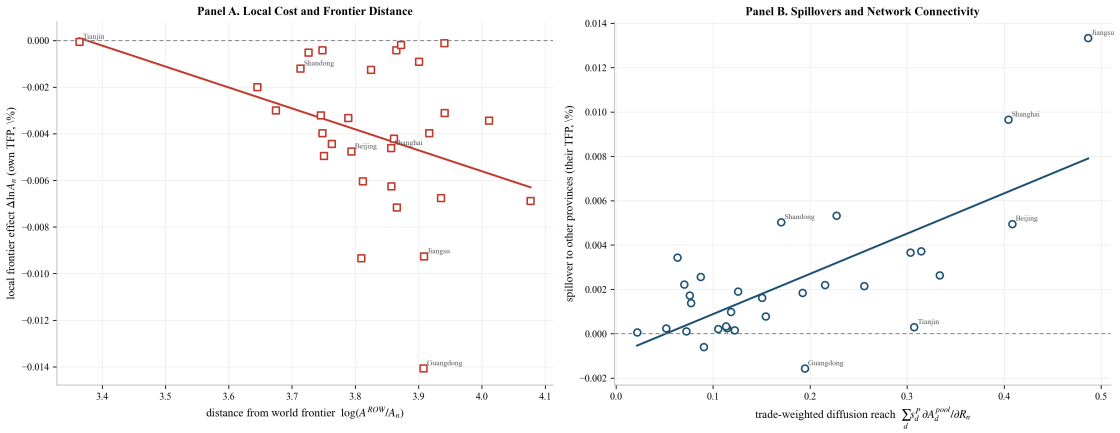
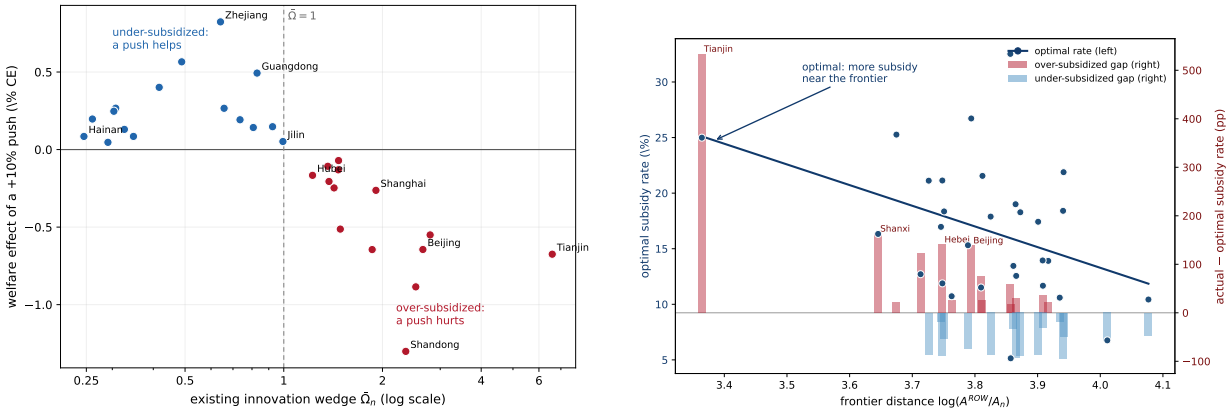



Figure 12: Local Reallocation Costs and External Diffusion Benefits

*Notes:* Each point raises one province's wedge by 10 percent from the undistorted allocation $\bar{\Omega}_n = 1$ and re-solves the stationary equilibrium. Panel A reports the province's own frontier response against distance from the world frontier. Panel B reports the production-labor-weighted frontier response of the other provinces against trade-weighted diffusion reach. Lines are OLS fits.

centives lower productivity in provinces with calibrated wedges well above the social optimum. The welfare-improving reform compresses these extremes while preserving moderate spatial variation that reflects each province's diffusion reach.




A. Marginal welfare effect

B. Optimal rates and calibrated-policy gaps

Figure 13: Welfare Incidence and the Optimal Subsidy Schedule

*Notes:* Panel A reports the change in China-wide consumption-equivalent worker welfare when one province's innovation wedge rises by 10 percent. Its horizontal axis is the calibrated wedge on a log scale, and the dashed line marks $\bar{\Omega}_n = 1$. Panel B plots welfare-optimal implicit subsidy rates against distance from the world frontier. Its bars report the calibrated implicit subsidy rate minus the welfare-optimal rate in percentage points, and the solid line is an OLS fit through the optimal rates.

# 7 Concluding Remarks

The Pigouvian case for innovation subsidies is correct on its own terms. Private R&D returns do fall short of social returns. It is incomplete in an economy where firms grow through absorption as well as invention, because a subsidy that closes the private-social gap also moves researchers, and the move has a price. With the local research stock fixed, raising the private payoff to frontier work draws effort away from absorption. The social cost of that reallocation is proportional to the local imitable pool, which a city's trade openness and the productivity of its partners determine. A policy concentrated in knowledge-rich, well-connected cities therefore imposes the largest absorption shortfall where absorption is worth most. A subsidy that weighs these forces is mild and jointly determined by the local reallocation cost and the location's diffusion reach.

We establish this mechanism empirically using an interacted difference-in-differences design estimated on outcome-specific samples drawn from a 3.25 million observation firm-year panel across 371 Chinese cities over 2000–2013. Innovation subsidies raise patenting on average but crowd out imitation, and both effects vary sharply with the local knowledge environment. In cities with greater access to external knowledge, the imitation crowd-out is deeper and the net effect on TFP turns negative; the sign reversal occurs only at approximately the 97th–98th percentile of the knowledge-access distribution, affecting roughly the top 2 percent of firms. Two independent instruments confirm that this gradient reflects a structural causal relationship: the first exploits exogenous connectivity gains from high-speed rail expansion along pre-existing slow-rail corridors, and the second constructs a supply-push shift-share measure of import-embodied technology inflows. The structural model translates these estimates into economy-wide welfare accounts. The calibrated innovation-payoff wedge varies by a factor of more than twenty across thirty Chinese provinces. It correlates strongly with innovation-policy intensity and is essentially uncorrelated with the raw subsidy receipt rate, so it measures the qualitative orientation of local policy and not its fiscal scale.

Reallocating the subsidy from the calibrated wedge to the welfare-optimal schedule raises aggregate manufacturing TFP by 11.26 percent and worker welfare by 4.25 percent. The optimal schedule balances frontier-related local losses against network-based diffusion benefits, and the welfare loss it recovers comes from current policy placing innovation pressure on locations where the opportunity cost in foregone absorption is high. A 10 percent reduction in bilateral trade costs raises aggregate TFP by a further 1.03 percent through a separate margin: cheaper trade expands the imitable pool while leaving the research stock where it is. The two reforms are complements.

Three implications follow. First, the social cost of an innovation subsidy depends on the local knowledge environment and not only on the gap between private and social R&D returns. China's experience illustrates both sides of this principle: subsidies raised TFP in knowledge-poor, isolated cities where the Pigouvian correction dominated, yet reduced it in knowledge-rich, well-connected cities where crowd-out of imitation outweighed the innovation gain. The same instrument can simultaneously correct a market failure in one location and deepen a misallocation in another. Second, China's aggregate innovation push succeeded in expanding frontier research capacity in its less-connected interior, yet the spatial distribution of subsidies was not aligned with the local knowledge environment. Knowledge-rich cities, which face the steepest imitation crowd-out, carried the highest calibrated innovation wedge, and the welfare cost of that mismatch is large relative to what a spatially differentiated policy could have achieved. Third, trade integration raises the stakes of subsidy design. Lower trade costs expand the imitable pool in every location, increasing the social return to imitation and therefore the welfare cost of any policy that tilts researchers excessively toward innovation. China's rapid trade integration over this period made the spatial misallocation of subsidies progressively more consequential, a dynamic that other rapidly integrating economies should anticipate.

More broadly, our findings identify two testable conditions under which the social cost of innovation subsidies is large: the local imitable pool must be rich, and policy must already tilt researchers toward innovation beyond the undistorted optimum. These conditions provide a portable framework for assessing when innovation-targeted industrial policy will impose spatial misallocation rather than correct a market failure. For economies with regionally heterogeneous knowledge endowments and expanding trade links, a description that fits much of East Asia, South Asia, and the emerging economies of Latin America and sub-Saharan Africa, this distinction is central to the design of technology policy.

# Online Appendix

*The Misallocated Innovation Subsidies in Space*

## Contents


**1 Introduction** **2**

**2 Empirical Evidence** **6**
2.1 Data . . . 6
2.2 Measurements, Facts, and Validations . . . 7
2.2.1 Industrial Policy in Space . . . 8
2.2.2 Knowledge Diffusion in Space . . . 11
2.2.3 General Knowledge Exposure . . . 16
2.2.4 High-Speed Rail and Knowledge Access . . . 18
2.2.5 Import-Embodied Technology Exposure . . . 18
2.2.6 Causal Identification . . . 19

**A Data Details** **40**
A.1 Firm-Level Production Data (ASIEC) . . . 40
A.2 Patent Records (CNIPA) . . . 40
A.3 Industrial Policy Documents . . . 40
A.4 City-Level Yearbooks and Trade Data . . . 41
A.5 Summary Statistics . . . 41
A.6 City-Level Patent Panel . . . 41

**B Main regression tables** **42**
B.1 Industrial policy in space . . . 42
B.2 Imitation measure: validation against patent citation network . . . 42

**C City-Level Innovation Composition: Supplementary Results** **43**

**D Measurement Details** **49**
D.1 TFP Estimation . . . 49
D.2 Technology Innovation Index . . . 49
D.3 Accessible Knowledge and Road-Cost Proxy . . . 49

**E Instrument Construction Details** **50**
E.1 HSR Planned-Network Shift-Share Instrument . . . 50
E.2 Historical Railway × Hub Knowledge Instrument . . . 51
E.3 Technology-Weighted Supply-Push Bartik Instrument . . . 51

**F Model Derivations and Equilibrium** **51**
F.1 Knowledge Diffusion and the Imitable Pool . . . 51
F.2 Frontier Draws and the Exact Scale Recursion . . . 55

**G Model Calibration and Counterfactual Algorithms** **59**
G.1 Province Innovation and Imitation Levels . . . 59
G.2 Reduced-Form Moments Used by the Model . . . 60
G.3 Common-Unit Normalization . . . 61
G.4 Baseline Equilibrium Inversion . . . 62
G.5 Counterfactual Solver . . . 63
G.6 Reported Outcomes . . . 64

**H Robustness Checks** **65**

**I Measurement of Breakthrough and Imitative Patents** **66**

**J Iceberg Trade Costs and Accessible Knowledge** **67**
J.1 Measuring Iceberg Trade Costs . . . 68
J.2 Measuring Accessible Knowledge . . . 68

**K Patent-Based Knowledge Diffusion** **70**
K.1 Patent-Level Technological Similarity . . . 70
K.2 City–City Technological Similarity Measures . . . 71
K.3 Citation-Based Measures of Knowledge Diffusion . . . 71
K.4 City-Level Aggregation of Residual-Based Imitation . . . 71

# A Data Details

## A.1 Firm-Level Production Data (ASIEC)

At the firm level, we observe a comprehensive cross-section of enterprise activities through the Annual Survey of Industrial Enterprises in China (ASIEC). The survey covers all industrial enterprises above the relevant scale threshold.[5] ASIEC is the most detailed enterprise census available for China, covering balance-sheet items, income-statement flows, and employee characteristics for each firm. Critically, ASIEC records government subsidy income as a separate income-statement line item, allowing us to observe the amount of subsidies received by each firm in each year.

To construct our analysis sample, we (i) drop firm-years with non-positive output, value-added, or employment; (ii) drop observations for which the capital-to-labour ratio or the intermediate-input-to-output ratio lies below the 1st or above the 99th percentile of the cross-sectional distribution in each year; (iii) require each firm to appear in at least two consecutive years. After these restrictions the baseline sample contains approximately 3.5 million firm-year observations.

## A.2 Patent Records (CNIPA)

The patent data are provided by the China National Intellectual Property Administration (CNIPA) and cover all patent applications filed in China between 1985 and 2018 (over 20 million records). For each application we observe: the application and grant dates, the IPC technology classification (up to the 4-digit subclass level), the legal status, the list of applicants with their registered addresses, and the forward and backward citation lists. We match patents to ASIEC firms using a three-step algorithm: (i) exact match on the enterprise registration code where available; (ii) probabilistic fuzzy match on firm name; (iii) geographic filter requiring the patent's registered city to match the firm's recorded city. Matched patents are aggregated to the firm-year level to construct patenting counts by type (invention, utility model, design), technology-class diversity indices, and citation-based quality measures.

## A.3 Industrial Policy Documents

To supplement the ASIEC subsidy-income variable, we collected the full text of all subsidy-related policy documents issued by municipal, provincial, and central-government bodies between 2000 and 2015. Documents were obtained from official government portals and the Peking University Law Database. We used keyword filtering to identify documents pertaining to technology subsidies, R&D support, innovation grants, and industrial upgrading programmes. For each document we recorded the issuing authority, the target firm or industry, the stated policy objective, and the grant amount where specified. This corpus allows us to construct subsidy-event indicators at the firm level (used to define "If Subsidy" in the empirical specifications) and to classify subsidies by their stated purpose.

[5] The definition of "above-scale" varies by survey year. From 1998 to 2006 the threshold is annual main-business revenue exceeding RMB 5 million; from 2007 to 2010 it rises to RMB 5 million for all enterprises; and from 2011 to 2013 it is RMB 20 million. We construct a consistent panel by applying a uniform deflator and dropping observations that do not satisfy the pre-2011 threshold throughout.

### A.4 City-Level Yearbooks and Trade Data

We compiled statistical yearbook data for all provinces, cities, and counties in China, covering aggregate GDP and its components, fiscal revenues and expenditures, fixed-asset investment, population, employment by sector, educational enrolment, and infrastructure investment. High-speed rail data—recording route-kilometres in operation by city and year—are taken from Ministry of Transport annual reports and cross-checked against press releases and station-opening announcements.

For the import-based knowledge instrument, customs transaction records from China's General Administration of Customs provide HS-6 product–source-country–destination-city trade flows at annual frequency from 1997 onward. We use these to construct city-level import shares by product and source country at the 2000 baseline, and to measure annual changes in source-country exports to comparison countries (Australia, Germany, Japan, New Zealand, Spain, Switzerland, the United Kingdom, and the United States).

### A.5 Summary Statistics

Table 10 reports summary statistics for the main variables used in the empirical analysis. The stacked panel contains 3,246,795 firm-year observations corresponding to 619,718 unique firms across 371 prefectural-level cities over 2000–2013; the table reports the non-missing ACF-TFP sample of 2,617,487 observations. Three features of the data are worth highlighting. First, innovation activity is highly concentrated: 95.6% of observations have a zero technology-innovation index, reflecting that patenting is dominated by a small fraction of technologically sophisticated firms. Second, subsidy receipt is selective but widespread: 11.5% of firm-year observations record positive government subsidy income in a given year. Third, there is substantial within-firm variation in TFP, suggesting that productivity growth rather than selection drives much of the aggregate dynamics. Aggregate subsidy disbursements grew nearly four-fold over the sample period, from RMB 31.6 billion in 2000 to RMB 140.9 billion in 2013, consistent with the expanding scope of China's innovation-policy agenda.

Table 11 documents the spatial distribution of subsidy rates, productivity, innovation, and imitation across cities. The table ranks cities rather than comparing raw means, with rank 1 denoting the highest value. Among the ten cities with the highest subsidy rates, the average TFP rank is 166, the average technology-innovation rank is 71, and the average imitation rank is 144. Among the ten cities with the lowest subsidy rates, the corresponding ranks are 53, 187, and 105. Thus high-subsidy cities innovate more but imitate less, while low-subsidy cities have higher TFP and stronger imitation. This contrast motivates the model's central mechanism: innovation subsidies can redirect scarce research effort toward own innovation and away from absorbing external knowledge.

### A.6 City-Level Patent Panel

Table 12 reports summary statistics for the city-year panel variables used in the city-level analysis in Section 3.1. The sample covers 287 prefectural-level cities over 2000–2013. *CumIP* denotes the cumulative count of industrial policy documents targeting each city–industry pair, standardised to have unit standard deviation. Mimic rate and BK rate are defined as the share of all city-year

Table 10: Summary Statistics

| Variable | $N$ | Mean | Std. Dev. | P25 | P50 | P75 |
|---|---|---|---|---|---|---|
| *A. Firm-Level Outcome Variables* | | | | | | |
| TFP (log, ACF estimator) | 2,617,487 | 3.446 | 1.028 | 2.809 | 3.420 | 4.082 |
| Technology Innovation Index ($\widetilde{Q}$) | 2,617,487 | 0.061 | 0.372 | 0.000 | 0.000 | 0.000 |
| Imitation Measure (*Immi*) | 2,378,931 | 0.000 | 0.596 | −0.292 | 0.014 | 0.324 |
| R&D Intensity (log, RMB thousands) | 1,303,538 | 5.920 | 2.483 | 4.578 | 6.110 | 7.546 |
| *B. Industrial Policy Variables* | | | | | | |
| Subsidy Indicator (0/1) | 2,617,487 | 0.115 | 0.320 | 0.000 | 0.000 | 0.000 |
| Subsidy Amount (log, RMB thousands) | 2,617,487 | 0.639 | 1.913 | 0.000 | 0.000 | 0.000 |
| *C. Firm Characteristics* | | | | | | |
| Firm Age (log years) | 2,616,307 | 2.060 | 0.793 | 1.609 | 2.079 | 2.565 |
| Exporter Indicator (0/1) | 2,617,487 | 0.240 | 0.427 | 0.000 | 0.000 | 0.000 |
| Log Employment | 2,617,487 | 5.034 | 1.087 | 4.263 | 5.017 | 5.724 |
| Log Capital Stock | 2,617,487 | 7.534 | 1.706 | 6.328 | 7.479 | 8.604 |
| Log Gross Output | 2,497,647 | 9.715 | 1.216 | 8.827 | 9.546 | 10.413 |

*Notes:* The table reports summary statistics for the main variables used in the empirical analysis. The sample covers 2,617,487 firm-year observations from the regression sample, covering 371 prefectural-level cities. TFP is estimated using the Ackerberg, Caves and Frazer (2015) proxy-variable method. The Technology Innovation Index $\widetilde{Q}$ is right-skewed: 95.6% of observations are zero (no patenting). The Subsidy Indicator equals one if the firm records positive government subsidy income. Subsidy Amount is measured in log RMB thousands and set to zero for non-recipients; the mean across all observations is 0.639 (Std. Dev. 1.913). Firm age is in log years; the median implied age is 8 years.

patents classified as imitative or breakthrough, respectively; the BK/Mimic ratio is in logs. All patent-based variables are winsorised at the 1st and 99th percentiles before first-differencing.

# B Main regression tables

## B.1 Industrial policy in space

Table 13 presents the city–year panel regressions that underlie the marginal-effect plots in Figure 2 of the main text. Columns (1)–(2) regress average firm TFP on city-level R&D subsidy totals and firm-level subsidy indicators, respectively, with city and year fixed effects and a rich set of city controls. Columns (3)–(4) add the interaction of each subsidy measure with the demeaned accessible-knowledge index; the interaction slope $\hat{\beta}_2 = -0.010^{***}$ (column 3) is the coefficient plotted in Figure 2, Panel A. Columns (5)–(6) replace accessible knowledge with the demeaned road-cost proxy; the interaction slope $\hat{\beta}_2 = +0.019^{***}$ (column 5) underlies Panel B.

## B.2 Imitation measure: validation against patent citation network

Table 14 validates the residual-based imitation measure $Immi_{ict}$ against six patent-citation outcomes from the CNIPA network. All six coefficients are positive and significant at the 1% level, confirming that the measure captures economically meaningful knowledge absorption from outside the firm's city.

# C City-Level Innovation Composition: Supplementary Results

Tables 15 and 16 report robustness checks for the city-level first-difference estimates in Table 2 (Section 3.1). Table 15 varies the specification one dimension at a time: Panel A replaces contemporaneous cumulative IP counts with a one-year lag; Panel B uses alternative patent-quality measures as the dependent variable; Panel C restricts the sample to cities with at least 50 patents in the base year. Table 16 presents additional checks including pre-trend tests, placebo IP assignments, and a balanced-panel restriction.

Table 11: Spatial Distribution of Industrial Policy, Innovation, and Productivity

| City / Region | Obs. | Subsidy Rate | Rank | TFP Rank | Tech Innov. Rank | Imitation Rank |
|---|---|---|---|---|---|---|
| *A. Ten Cities with Highest Subsidy Rates* | | | | | | |
| Xining (Qinghai) | 1,290 | 0.316 | 1 | 158 | 172 | 126 |
| Qujing (Yunnan) | 2,556 | 0.288 | 2 | 67 | 190 | 204 |
| Quzhou (Zhejiang) | 6,983 | 0.275 | 3 | 137 | 68 | 158 |
| Urumqi (Xinjiang) | 3,129 | 0.275 | 4 | 202 | 30 | 186 |
| Jinhua (Zhejiang) | 29,215 | 0.269 | 5 | 215 | 83 | 153 |
| Kunming | 6,251 | 0.265 | 6 | 196 | 35 | 141 |
| Tongling (Anhui) | 1,662 | 0.262 | 7 | 200 | 2 | 105 |
| Lishui (Zhejiang) | 7,392 | 0.260 | 8 | 183 | 49 | 37 |
| Ningbo | 76,743 | 0.258 | 9 | 205 | 28 | 93 |
| Haikou | 1,668 | 0.257 | 10 | 97 | 53 | 238 |
| *B. Ten Cities with Lowest Subsidy Rates (min. 1,000 obs.)* | | | | | | |
| Dezhou (Shandong) | 18,267 | 0.021 | 244 | 50 | 184 | 56 |
| Pingxiang (Jiangxi) | 3,708 | 0.021 | 243 | 39 | 226 | 233 |
| Zhoukou (Henan) | 4,745 | 0.022 | 242 | 59 | 240 | 210 |
| Kaifeng (Henan) | 5,252 | 0.028 | 241 | 96 | 235 | 96 |
| Xiantao (Hubei) | 2,336 | 0.029 | 240 | 20 | 178 | 1 |
| Puyang (Henan) | 3,739 | 0.034 | 239 | 13 | 88 | 75 |
| Zhumadian (Henan) | 5,101 | 0.035 | 238 | 98 | 230 | 142 |
| Weifang | 32,222 | 0.035 | 237 | 48 | 116 | 84 |
| Heze (Shandong) | 10,558 | 0.037 | 236 | 93 | 196 | 29 |
| Zaozhuang | 9,973 | 0.037 | 235 | 10 | 177 | 119 |
| *C. Obs.-Weighted Average Rank by Coastal vs. Inland Status* | | | | | | |
| Coastal cities (37 cities) | 802,569 | 0.140 | – | 165 | 56 | 106 |
| Inland cities (334 cities) | 1,814,918 | 0.104 | – | 125 | 92 | 116 |
| Full sample | 2,617,487 | 0.115 | – | 138 | 81 | 113 |

*Notes:* The table summarises the spatial distribution of innovation subsidies and firm-level outcomes across prefectural cities, years 2000–2007 and 2011–2013 (regression sample, $N$ =2,617,487). Subsidy Rate is the share of firm-year observations with positive government subsidy income. TFP Rank, Innov. Rank, and Imitation Rank are each city's rank among all 244 qualifying cities (minimum 1,000 observations) based on city-mean log-TFP (Ackerberg, Caves and Frazer 2015), city-mean technology-innovation index $\widetilde{Q}$, and city-mean imitation measure *Immi*, respectively; rank 1 denotes the highest value. Panel C reports the observation-weighted average rank for coastal and inland cities. Panels A and B reveal a striking pattern: high-subsidy cities are predominantly inland, resource-rich, or peripherally located, and exhibit systematically lower technology innovation and more negative imitation scores, consistent with the model's prediction that subsidies are misallocated toward locations where the imitable knowledge pool is shallow. Panel C confirms the aggregate coastal–inland contrast.

Table 12: Summary Statistics: Key Regression Variables (247 Cities, 2000–2013, $N = 3,458$)

| Category | Variable | $N$ | Mean | SD | P10 | Median | P90 |
|---|---|---|---|---|---|---|---|
| *Regressor* | ln(Cum. IP docs+1) | 3,458 | 0.315 | 0.619 | 0.000 | 0.000 | 1.099 |
| | ln(Annual IP docs+1) | 3,458 | 0.098 | 0.301 | 0.000 | 0.000 | 0.693 |
| *Composition outcomes* | Mimic share of classified | 1,599 | 0.524 | 0.405 | 0.000 | 0.400 | 1.000 |
| | ln(BK/Mimic) | 3,458 | 0.055 | 0.890 | −1.099 | 0.000 | 1.344 |
| *Pool-rate outcomes* | BK rate ($n_{\text{BK}}/n_{\text{BK-elig}}$) | 2,542 | 0.103 | 0.166 | 0.000 | 0.000 | 0.357 |
| | Mimic rate ($n_{\text{Mimic}}/n_{\text{Mimic-elig}}$) | 2,542 | 0.056 | 0.092 | 0.000 | 0.024 | 0.143 |
| *Broad non-innovation* | Broad non-innov. share | 1,651 | 0.549 | 0.404 | 0.000 | 0.500 | 1.000 |
| *Heterogeneity* | Mimic share (other entities) | 1,490 | 0.508 | 0.418 | 0.000 | 0.364 | 1.000 |
| *Scale (reference)* | ln(BK+1) | 3,458 | 0.771 | 1.431 | 0.000 | 0.000 | 2.996 |
| | ln(Mimic+1) | 3,458 | 0.718 | 1.158 | 0.000 | 0.000 | 2.485 |

*Notes:* Balanced city-year panel of 247 prefecture-level cities, 2000–2013. Missing patent observations filled with zero. All continuous variables winsorized at the 1st–99th percentile. China-domestic (ctry) top-decile classification throughout. *BK rate* $= n_{\text{BK}}/n_{\text{total}}$; *Mimic rate* $= n_{\text{Mimic}}/n_{\text{total}}$; *ln(BK/Mimic)* $= \ln(n_{\text{BK}}+1) - \ln(n_{\text{Mimic}}+1)$. IP stock is the cumulative count of IP-related policy documents issued by the city government through year $t$.

Table 13: The Geography of Industrial Policies

| | *Dependent Variable: City Average TFP* | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| RD Subsidy City Total | 0.007* | | 0.010*** | | 0.015*** | |
| | (0.003) | | (0.003) | | (0.004) | |
| Subsidy | | 0.002* | | 0.001 | | 0.001 |
| | | (0.001) | | (0.001) | | (0.001) |
| RD Subsidy City Total × Accessible Knowledge | | | −0.018*** | | | |
| | | | (0.007) | | | |
| Subsidy × Accessible Knowledge | | | | −0.005** | | |
| | | | | (0.002) | | |
| RD Subsidy City Total × Iceberg Cost | | | | | 0.019*** | |
| | | | | | (0.005) | |
| Subsidy × Iceberg Cost | | | | | | 0.001** |
| | | | | | | (0.001) |
| Observations | 2,365,061 | 1,871,471 | 2,365,061 | 1,870,938 | 2,365,061 | 1,870,938 |
| $R^2$ | 0.828 | 0.815 | 0.839 | 0.833 | 0.833 | 0.816 |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| City FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| City Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

*Notes:* The table reports OLS estimates of the relationship between innovation subsidies and city-average firm TFP. The sample is the universe of industrial firms in the ASIEC. Odd columns use city-level R&D subsidy totals; even columns use firm-level subsidy receipts. Columns (3)–(4) interact the subsidy variable with city-level accessible knowledge; columns (5)–(6) interact with the road-cost proxy labelled Iceberg Cost. City controls include log population, log per-capita GDP, and the industrial output share of manufacturing. Standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% levels.

Table 14: Imitation and Patent-Based Knowledge Diffusion

| | (1) Citations made per patent | (2) Similar-patent share | (3) Similarity intensity |
|---|---|---|---|
| City imitation index | 0.0052*** | 0.0040*** | 0.0033*** |
| | (0.0013) | (0.0003) | (0.0002) |
| Observations | 26,140 | 1,075,193 | 1,037,522 |
| Origin-city FE | ✓ | ✓ | ✓ |
| Destination-city FE | ✓ | ✓ | ✓ |
| Year FE | ✓ | ✓ | ✓ |

*Notes:* The table validates the firm-level imitation residual against two independent parts of the patent record. The unit of observation is a city-pair–year, and the regressor is the city-year mean of $Immi_{ict}$ in the absorbing city, the index used by the baseline specifications. *Column (1):* the outcome is the number of citations the citing city's patents make to the other city's patents at a five-year horizon, divided by the citing city's own patent count. Scaling by that count measures citation propensity and holds patent volume fixed. Entering the cited city's index alongside leaves it indistinguishable from zero ($-0.001$, s.e. 0.001), so the relation runs through the city doing the citing. *Columns (2)–(3):* the outcome is the number of the later city's patents similar to the earlier city's patents as a share of all pairs (col. 2), and the log mean similarity among the pairs that clear the 90th-percentile threshold (col. 3). Similarity is cosine similarity in patent-text space. Both are free of patent volume. All three coefficients keep their sign, significance and order of magnitude on an imitation index built without interpolating gaps in a firm's series (0.009, 0.001 and 0.003, each significant at the 1% level). Standard errors are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% levels.

Table 15: Robustness: City-Level Innovation Composition

| | (1) | (2) | (3) |
|---|---|---|---|
| | *Composition shift (FD)* | | *Innovation scale (OLS)* |
| | $\Delta$ Mimic rate | $\Delta$ BK rate | ln(BK/Mimic) |
| *Panel A: Province× Year FE* | | | |
| $\Delta \ln(\text{CumIP}+1)$ | $-0.029^{*}$ | $0.037^{**}$ | $0.218^{***}$ |
| | (0.016) | (0.017) | (0.057) |
| Observations | 1,793 | 1,793 | 3,206 |
| *Panel B: Control for total patents ($+\Delta \ln n_{total}$ / $+\ln n_{total}$)* | | | |
| $\Delta \ln(\text{CumIP}+1)$ | $-0.021^{*}$ | $0.032^{**}$ | $0.336^{***}$ |
| | (0.012) | (0.016) | (0.050) |
| Observations | 2,056 | 2,056 | 3,458 |
| *Panel C: 1%–99% winsorization* | | | |
| $\Delta \ln(\text{CumIP}+1)$ | $-0.024^{*}$ | $0.026^{*}$ | $0.317^{***}$ |
| | (0.014) | (0.013) | (0.051) |
| Observations | 2,056 | 2,056 | 3,458 |
| City FE | *(by diff.)* | | ✓ |
| Year FE | *(see panel)* | | ✓ |

*Notes:* Robustness checks replicating Table 2. Cols. (1)–(2): first-difference specification; city FE absorbed by differencing. Col. (3): OLS levels with city and year FE. *Panel A*: province×year FE absorb regional policy cycles. *Panel B*: controls for total patent volume ($\Delta \ln n_{ct}^{\text{total}}$ in FD cols; $\ln n_{ct}^{\text{total}}$ in OLS col). *Panel C*: 1%–99% winsorization of outcome and treatment variables. Standard errors are reported in parentheses. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, and 10% levels.

Table 16: Additional Robustness: City-Level Innovation Composition

| | (1) | (2) | (3) |
|---|---|---|---|
| | *Composition shift (FD)* | | *Innovation scale (OLS)* |
| | $\Delta$ Mimic rate | $\Delta$ BK rate | ln(BK/Mimic) |
| *Panel D: Alternative treatment — current-year IP ($\Delta \ln(IP_{ct}+1)$ / $\ln(IP_{ct}+1)$)* | | | |
| $\Delta \ln(\text{CumIP}+1)$ | $-0.009^{*}$ | $0.017^{*}$ | $0.216^{***}$ |
| | (0.005) | (0.009) | (0.061) |
| Observations | 2,056 | 2,056 | 3,458 |
| *Panel E: Early sub-period (2000–2006)* | | | |
| $\Delta \ln(\text{CumIP}+1)$ | $-0.041$ | 0.003 | $-0.158^{*}$ |
| | (0.026) | (0.003) | (0.081) |
| Observations | 1,012 | 1,012 | 1,729 |
| *Panel F: Late sub-period (2007–2013)* | | | |
| $\Delta \ln(\text{CumIP}+1)$ | $-0.011$ | $0.047^{**}$ | $0.394^{***}$ |
| | (0.013) | (0.023) | (0.085) |
| Observations | 1,044 | 1,044 | 1,729 |
| *Panel G: Placebo — within-year city permutation (500 iterations)* | | | |
| Actual coefficient | $-0.0201$ | $+0.0329$ | $+0.3172$ |
| Placebo $p$-value | 0.140 | $<0.05$ | $<0.01$ |
| City FE | *(by diff.)* | | ✓ |
| Year FE | ✓ | ✓ | ✓ |

*Notes:* Additional robustness checks replicating Table 2. Cols. (1)–(2): first-difference specification; city FE absorbed by differencing. Col. (3): OLS levels with city and year FE. *Panel D*: replaces cumulative IP stock $\text{CumIP}_{ct}$ with current-year IP document count $\text{IP}_{ct}$. *Panels E–F*: restrict to early (2000–2006) and late (2007–2013) sub-periods. *Panel G*: within each year, city IP stocks are randomly permuted; 500 iterations; placebo $p$-value = share of permutation draws with $|\hat{\beta}_{\text{placebo}}| \geq |\hat{\beta}_{\text{actual}}|$. Standard errors are reported in parentheses. $^{***}$, $^{**}$, $^{*}$ denote significance at the 1%, 5%, and 10% levels.

# D Measurement Details

## D.1 TFP Estimation

We estimate firm-level TFP using the two-stage proxy-variable estimator of Ackerberg, Caves and Frazer (2015), implemented separately for each two-digit CIC industry. In the first stage we regress log output on log capital, log employment, log intermediate inputs, and a flexible polynomial in log capital and log intermediate inputs (the proxy for unobserved productivity). In the second stage we recover the production-function coefficients by GMM, using lagged intermediate inputs as instruments for current productivity and assuming that TFP follows a first-order Markov process. We report robustness to the Olley and Pakes (1996) estimator in the robustness section of the appendix.

## D.2 Technology Innovation Index

Let $Z^{IU}_{icgt}$ denote the number of invention and utility-model patent applications filed by firm $i$ in city $c$, year $t$, and IPC technology group $g$. Let $Z^{\text{All}}_{ict}$ denote the firm's total number of invention, utility-model, and design patent applications. We define the Herfindahl-based technology-diversity component as

$$\mathcal{D}_{ict} = \begin{cases} 1 - \sum_{g} \left( \dfrac{Z^{IU}_{icgt}}{Z^{\text{All}}_{ict}} \right)^2, & Z^{\text{All}}_{ict} > 0, \\ 0, & Z^{\text{All}}_{ict} = 0. \end{cases}$$

The technology-innovation index used in the empirical regressions is

$$\widetilde{Q}_{ict} \equiv \pi_{ict} = \begin{cases} \ln\left( Z^{\text{All}}_{ict} \mathcal{D}_{ict} \right), & Z^{\text{All}}_{ict} \mathcal{D}_{ict} > 0, \\ 0, & Z^{\text{All}}_{ict} \mathcal{D}_{ict} = 0 \text{ or either component is unobserved.} \end{cases}$$

This is the transformation applied in the estimation sample: the logarithm is taken without adding one, and observations for which the logarithm is undefined are assigned zero. The index combines patenting volume with dispersion across technology fields.

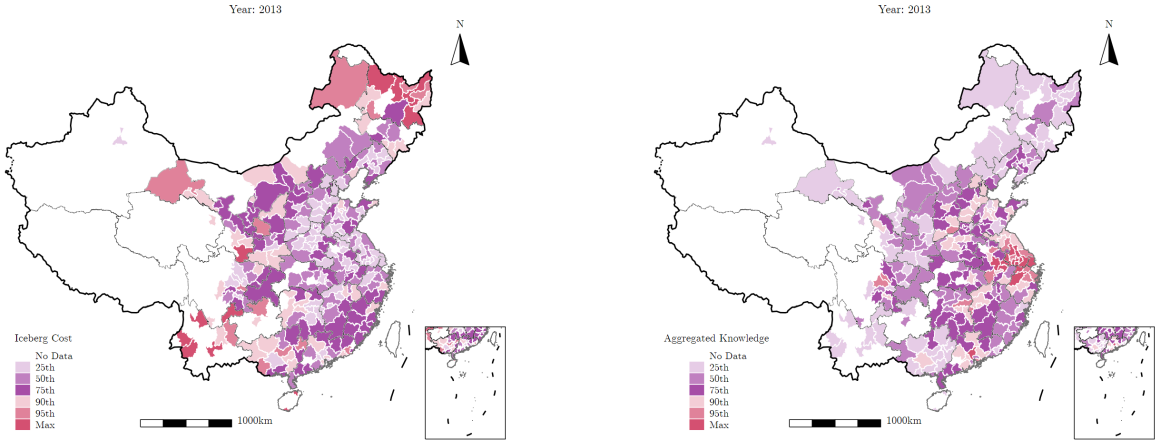
## D.3 Accessible Knowledge and Road-Cost Proxy

For each city pair $(c, d)$ and year $t$, let $r_{cdt}$ denote the bilateral road travel cost from the prefecture-pair road network matrix. For each origin city $c$, define $\mathcal{N}_{30}(c, t)$ as the thirty outside destination cities with the lowest values of $r_{cdt}$. We construct Accessible Knowledge as

$$A_{ct} = \ln\left( 1 + \sum_{d \in \mathcal{N}_{30}(c,t)} \frac{\exp(\bar{\omega}_{dt})}{r_{cdt}} \right),$$

where $\bar{\omega}_{dt}$ is the firm-average log-TFP in destination city $d$ and year $t$. The distance-decay exponent is therefore fixed at one in the implementation. We also report an "Iceberg Cost" variable in the tables, defined as $100^{-1} \sum_{d \in \mathcal{N}_{30}(c,t)} r_{cdt}$. This variable is a road-cost proxy for the cost of reaching nearby outside cities; it is not recovered as a residual from a gravity equation. The maps use the same road matrix, with a nearest-10 variant for visualisation.

Figure 14 maps both measures for 2013 across all prefectural cities.




A. Road-Cost Proxy (Iceberg Cost), 2013

B. Aggregate Knowledge, 2013

Figure 14: Knowledge accessibility and road costs across Chinese cities

*Notes:* Panel A maps the road-cost proxy (Iceberg Cost) for each prefectural city in 2013; darker shading indicates higher costs. Panel B maps accessible knowledge in 2013, constructed per equation (4). Both variables exhibit strong East–West gradients and are the primary spatial heterogeneity dimensions in the analysis.

# E Instrument Construction Details

This appendix describes the construction of both instrumental variables used in Section 2.2.6 and provides first-stage diagnostic statistics.

## E.1 HSR Planned-Network Shift-Share Instrument

*Step 1: Theoretical planned network.* China's HSR network was conceived in the 1990s to link provincial capitals through geographically efficient corridors. We operationalise this planning logic by constructing a theoretical network consisting of the great-circle segments connecting every pair of the 30 provincial capitals ($\binom{30}{2} = 435$ segments). City $c$ is assigned $\text{PlannedNet}_c = 1$ if its geographic centroid lies within 50 km of at least one such segment. This indicator is time-invariant and determined entirely by physical geography; it is orthogonal to post-2003 economic conditions by construction.

*Step 2: Hub classification.* We use pre-sample (2003) city-level average TFP, estimated using the Ackerberg, Caves and Frazer (2015) method, as the measure of baseline knowledge. Cities above the 80th percentile of this distribution are classified as *knowledge hubs* ($d \in \mathcal{H}$). The hub classification is frozen at the 2003 baseline and is not updated as the panel progresses.

*Step 3: Instrument construction.* The preferred shift-share instrument is:

$$Z_{ct} = \text{PlannedNet}_c \times \ln\left(1 + \sum_{d \in \mathcal{H}} \widetilde{K}_d \cdot \mathbf{1}[\text{HSR}_{dt} = 1]\right), \tag{25}$$

where $\widetilde{K}_d = \exp(\hat{\omega}_d^{2003})$ is hub $d$'s baseline TFP level and $\text{HSR}_{dt} = 1$ if city $d$ had positive route-kilometres of HSR in operation by year $t$. The inner term is the TFP-weighted cumulative knowledge stock of hub cities with active HSR service; the logarithm mitigates skewness; and the product with $\text{PlannedNet}_c$ generates the required cross-city variation.

### E.2 Historical Railway × Hub Knowledge Instrument

For the historical-railway IV, we replace the planned-network indicator with an indicator for the presence of a functioning railway in city $c$ in 1933, digitised from Republican-era transport maps. This indicator predates the sample period by 70 years and reflects geographical and strategic considerations of a different political era, making it plausibly orthogonal to contemporary economic conditions. The instrument is:

$$\mathrm{IV}_{c,t}^{\mathrm{HSR}} = \mathbf{1}[\mathrm{Railway}_{c,1933}] \times \Psi_t, \tag{26}$$

where $\Psi_t = \ln(1 + \sum_{d \in \mathcal{H}} \widetilde{K}_{d,0} \cdot \mathbf{1}[\mathrm{HSR}_{dt} = 1])$ is the year-specific component common to all cities. First-stage $F$-statistics for both instrument variants consistently exceed the Stock–Yogo threshold for strong instruments (see Table 6 and Table 7 in the main text for the preferred specifications).

### E.3 Technology-Weighted Supply-Push Bartik Instrument

Formal construction follows equations (7)–(8) in the main text. The baseline industrial composition shares $\{s_{c,k,t_0}\}$ are measured at year 2000 using city-industry output from ASIEC. HS-6 product-level shocks are mapped to two-digit CIC industries using the UN Comtrade–CIC concordance provided by the CEPII; for many-to-one mappings we split weights equally. Comparison countries for third-country exports are Australia, Germany, Japan, New Zealand, Spain, Switzerland, the United Kingdom, and the United States—the same set used by Ackerberg, Caves and Frazer (2015) in the ADH design.

## F Model Derivations and Equilibrium

This section supplies the derivations omitted from the compact model in Section 3 of the main text. We follow the order of the main text. Standard CES demand and Type-I extreme-value choice results are stated only when they are needed to close the equilibrium. The nonstandard diffusion, correlated-trade, frontier, and research-allocation blocks are derived explicitly.

### F.1 Knowledge Diffusion and the Imitable Pool

This subsection derives every object in equations (19) and (20). The timing is as follows. An idea is created in origin $i$ during date $s$. At the beginning of date $t-1$, it may already be known in several locations. During date $t-1$, trade may make it observable in destination $n$. At the beginning of date $t$, researchers in $n$ may learn it. Observation therefore precedes learning by one period.

**Idea cohorts.** Innovation in location $i$ at date $s$ creates a Poisson point process of ideas over intrinsic qualities $v > 0$ with intensity measure

$$\mathcal{I}_{i,s}(dv) = \theta v^{-\theta-1} R_{i,s}\, dv. \tag{27}$$

For any threshold $u > 0$, the expected number of cohort-$(i,s)$ ideas above $u$ is

$$\mathcal{I}_{i,s}((u,\infty)) = \int_u^\infty \theta v^{-\theta-1} R_{i,s}\, dv = R_{i,s} u^{-\theta}. \tag{28}$$

The last equality follows because $\int_u^\infty \theta v^{-\theta-1} dv = [-v^{-\theta}]_u^\infty = u^{-\theta}$.

For each destination $n$, let $\sigma_{in}^{s,t} \in [0,1]$ denote the share of cohort-$(i,s)$ ideas known in $n$ at the beginning of date $t$. An idea is known at birth in its origin and nowhere else:

$$\sigma_{ii}^{s,s} = 1, \qquad \sigma_{in}^{s,s} = 0 \quad (n \neq i). \tag{29}$$

**Route-level observation and learning.** Introduce a lower-tail cutoff $\bar{z} > 0$ and normalize quality by $x = v/\bar{z} \geq 1$. Recall that $\lambda_{nm,t-1}$ is importer $n$'s expenditure share on source $m$. Conditional on an idea being known in $m$, the probability that route $m \to n$ makes it observable during $t-1$ is

$$G_{nm}^{t-1}(v, \bar{z}) \equiv \lambda_{nm,t-1} \phi_{nm}(v/\bar{z}), \tag{30}$$

where $0 \leq \phi_{nm}(x) \leq 1$, $\phi'_{nm}(x) \geq 0$, and $\phi_{nm}(x) \to 1$ as $x \to \infty$. Conditional on becoming observable, the idea is learned at the beginning of $t$ with probability

$$S_{n,t} \psi_n(v/\bar{z}), \tag{31}$$

where $0 \leq S_{n,t} \leq 1$, $0 < \psi_n(x) \leq 1$, and $\psi_n(x) \to d \in (0,1]$ as $x \to \infty$. For a fixed idea, we assume that knowledge indicators across source locations and route-level observation events are conditionally independent. These assumptions give the quality-dependent microfoundation; the BGP implementation below is its upper-tail limit.

**Proposition 1 (Cross-route observability)** *Fix a cohort-$(i,s)$ idea of quality $v \geq \bar{z}$. The probability that the idea becomes observable in $n$ through at least one route during $t-1$ is*

$$G_{in}^{s,t-1}(v, \bar{z}) = 1 - \prod_{m \neq n} \left[1 - \sigma_{im}^{s,t-1} G_{nm}^{t-1}(v, \bar{z})\right]. \tag{32}$$

*As $\bar{z} \to 0$, this probability converges to*

$$G_{in}^{s,t-1} = 1 - \prod_{m \neq n} \left(1 - \sigma_{im}^{s,t-1} \lambda_{nm,t-1}\right). \tag{33}$$

**Proof.** For each $m \neq n$, define

$$D_{im}^{s,t-1} = \{\text{the idea is known in } m \text{ at the beginning of } t-1\}$$

and

$$E_{nm}^{t-1} = \{\text{the idea is observed in } n \text{ through } m \to n \text{ during } t-1\}.$$

By the definition of $\sigma$, $\Pr(D_{im}^{s,t-1}) = \sigma_{im}^{s,t-1}$. By equation (30),

$$\Pr(E_{nm}^{t-1} \mid D_{im}^{s,t-1}) = G_{nm}^{t-1}(v, \bar{z}).$$

Hence

$$\Pr(D_{im}^{s,t-1} \cap E_{nm}^{t-1}) = \sigma_{im}^{s,t-1} G_{nm}^{t-1}(v, \bar{z}).$$

The probability that route $m \to n$ does not expose the idea is therefore $1 - \sigma_{im}^{s,t-1} G_{nm}^{t-1}(v, \bar{z})$. Conditional independence across routes implies that the probability of no exposure through any $m \neq n$ is the product in equation (32). Taking its complement proves that equation.

For fixed $v > 0$, $v/\bar{z} \to \infty$ as $\bar{z} \to 0$. Thus $\phi_{nm}(v/\bar{z}) \to 1$ and $G_{nm}^{t-1}(v, \bar{z}) \to \lambda_{nm,t-1}$. Continuity of the finite product then gives equation (33). ■

**Proposition 2 (Newly learned ideas)** *Let $\mu^{\text{adopt}}_{in,s,t}((u,\infty);\bar{z})$ be the expected number of cohort-$(i,s)$ ideas with intrinsic quality above $u$ that are newly learned in $n$ at the beginning of $t$. For $u \geq \bar{z}$,*

$$\mu^{\text{adopt}}_{in,s,t}((u,\infty);\bar{z}) = \int_u^\infty \left(1-\sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1}(v,\bar{z}) S_{n,t}\psi_n(v/\bar{z}) \times \theta v^{-\theta-1} R_{i,s}\, dv. \tag{34}$$

*In the upper-tail limit,*

$$\lim_{\bar{z}\to 0} \mu^{\text{adopt}}_{in,s,t}((u,\infty);\bar{z}) = \left(1-\sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} S_{n,t} d R_{i,s} u^{-\theta}. \tag{35}$$

**Proof.** An idea of quality $v$ is newly learned at $t$ if and only if three events occur: it was not already known in $n$, it becomes observable during $t-1$, and researchers successfully learn it at $t$. Their joint probability is

$$\left(1-\sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1}(v,\bar{z}) S_{n,t}\psi_n(v/\bar{z}).$$

Thinning the Poisson point process (27) by this quality-dependent probability multiplies its intensity by the same probability. Integrating the thinned intensity over $(u,\infty)$ gives (34).

For each fixed $v \geq u > 0$, Proposition 1 and the assumptions on $\psi_n$ imply

$$G_{in}^{s,t-1}(v,\bar{z})\psi_n(v/\bar{z}) \longrightarrow G_{in}^{s,t-1} d.$$

The remaining probability factors lie in $[0,1]$, so the absolute value of the integrand is bounded by $\theta v^{-\theta-1} R_{i,s}$, whose integral over $(u,\infty)$ is finite by (28). Dominated convergence therefore permits interchanging the limit and integral. Pulling the limiting probability outside the integral and using (28) gives (35). ■

**Learning effort and the imitable pool.** Local imitation effort is

$$e_{n,t} = a_n^M (L_{n,t}^M)^\nu. \tag{36}$$

Effort per available idea determines destination learning intensity:

$$S_{n,t} = \left(\frac{e_{n,t}}{A_{n,t}^{\text{pool}}}\right)^\eta, \qquad s_{n,t} \equiv d S_{n,t}, \qquad 0 \leq s_{n,t} \leq d \leq 1. \tag{37}$$

The bound is the unsaturated-learning condition: a location cannot learn more ideas than are available to it.

Before learning occurs, a cohort idea is in destination $n$'s imitable set if it is not yet known there and becomes observable during $t-1$. In the upper-tail system, the expected number of such cohort ideas above quality $u$ is

$$\left(1-\sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} R_{i,s} u^{-\theta}. \tag{38}$$

Summing across all earlier foreign cohorts gives

$$\mathbb{E}[\#\{\text{imitable ideas in } n \text{ at } t \text{ with quality} > u\}] = A_{n,t}^{\text{pool}} u^{-\theta}, \tag{39}$$

$$A_{n,t}^{\text{pool}} = \sum_{i\neq n}\sum_{s<t} \left(1-\sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} R_{i,s}. \tag{40}$$

This proves equation (19). The restriction $i \neq n$ prevents domestic innovation from being counted twice: it enters the domestic frontier directly and may seed other locations' pools, but it is not domestic imitation.

**Proposition 3 (Diffusion law of motion)** *The share of cohort-$(i,s)$ ideas known in $n$ evolves according to*

$$\sigma_{in}^{s,t} = \sigma_{in}^{s,t-1} + \left(1 - \sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} s_{n,t}, \tag{41}$$

*with initial conditions* (29).

**Proof.** By (28), the expected number of cohort ideas above $u$ known in $n$ at the beginning of $t-1$ is

$$\sigma_{in}^{s,t-1} R_{i,s} u^{-\theta}.$$

By Proposition 2 and $s_{n,t} = dS_{n,t}$, the expected newly learned flow is

$$\left(1 - \sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} s_{n,t} R_{i,s} u^{-\theta}.$$

Known ideas at $t$ equal the previously known stock plus this new flow:

$$\begin{aligned} \sigma_{in}^{s,t} R_{i,s} u^{-\theta} = {} & \sigma_{in}^{s,t-1} R_{i,s} u^{-\theta} \\ & + \left(1 - \sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} s_{n,t} R_{i,s} u^{-\theta}. \end{aligned}$$

Dividing throughout by the strictly positive $R_{i,s} u^{-\theta}$ gives (41). Together with (33) and (29), this equation recursively determines the complete diffusion history. ■

**From newly learned ideas to the imitation frontier.** Equations (39) and (37) imply that newly learned ideas form a Poisson point process with intrinsic-quality tail intensity

$$\Xi_{n,t}^{\text{im}} u^{-\theta}, \qquad \Xi_{n,t}^{\text{im}} \equiv s_{n,t} A_{n,t}^{\text{pool}}. \tag{42}$$

Each learned idea has a destination-specific applicability draw $\xi \sim$ Fréchet$(\vartheta)$, independent of intrinsic quality, with

$$\Pr(\xi \leq x) = \exp(-x^{-\vartheta}), \qquad \vartheta > \theta.$$

Let $T = \xi^{-\vartheta}$. Then $T \sim$ Exponential(1) and, setting $r = \theta/\vartheta \in (0,1)$,

$$\mathbb{E}[\xi^{\theta}] = \mathbb{E}[T^{-r}] = \int_0^{\infty} t^{-r} e^{-t}\, dt = \Gamma(1-r) = \Gamma\left(1 - \frac{\theta}{\vartheta}\right). \tag{43}$$

Conditional on an applicability draw $\xi$, an idea with intrinsic quality $v$ has effective quality $v\xi$. The expected number of newly learned ideas with effective quality above $q$ is therefore

$$\begin{aligned} \mathbb{E}[\#\{v\xi > q\}] &= \mathbb{E}_{\xi}\left[\Xi_{n,t}^{\text{im}} (q/\xi)^{-\theta}\right] \\ &= \Xi_{n,t}^{\text{im}} q^{-\theta} \mathbb{E}[\xi^{\theta}] \\ &= \Gamma\left(1 - \frac{\theta}{\vartheta}\right) \Xi_{n,t}^{\text{im}} q^{-\theta}. \end{aligned} \tag{44}$$

The Poisson void probability now gives the distribution of the best newly learned effective idea:

$$\begin{aligned}\Pr(q_{n,t}^{\text{im}} \le q) &= \Pr\{\#(v\xi > q) = 0\} \\ &= \exp\left[-\Gamma\left(1-\frac{\theta}{\vartheta}\right)\Xi_{n,t}^{\text{im}} q^{-\theta}\right]. \end{aligned} \tag{45}$$

Thus the imitation draw is Fréchet with frontier scale

$$\begin{aligned}A_{n,t}^{\text{im}} &= \Gamma\left(1-\frac{\theta}{\vartheta}\right)\Xi_{n,t}^{\text{im}} \\ &= \Gamma\left(1-\frac{\theta}{\vartheta}\right) s_{n,t} A_{n,t}^{\text{pool}} \\ &= \Gamma\left(1-\frac{\theta}{\vartheta}\right) d(a_n^M)^{\eta}(A_{n,t}^{\text{pool}})^{1-\eta}(L_{n,t}^M)^{\nu\eta}. \end{aligned} \tag{46}$$

This proves equation (20) and shows exactly where the $\Gamma(1-\theta/\vartheta)$ term comes from. It is the $\theta$th moment of the applicability draw, not an additional normalization.

### F.2 Frontier Draws and the Exact Scale Recursion

This subsection derives equation (21). All statements are conditional on the period-$t$ state, which includes the innovation intensity, the diffusion history, and the sets of ideas known before date $t$. The derivation uses exact distributional identities. In particular, the object $A_{n,t}$ is the scale appearing in the exponent of a Fréchet cumulative distribution function; it is not an approximation to expected productivity.

**A marked-Poisson maximum.** We first establish the common result used for both innovation and imitation.

**Lemma 4 (Maximum of marked Poisson ideas)** *Let intrinsic idea qualities form a Poisson point process on $(0,\infty)$ with intensity*

$$\theta\Xi u^{-\theta-1}\,du, \qquad \Xi \ge 0, \qquad \theta > 0. \tag{47}$$

*Attach to every idea an independent applicability mark $\xi$ with cumulative distribution function*

$$F_\xi(x) = \exp(-x^{-\vartheta}), \qquad x > 0, \qquad \vartheta > \theta. \tag{48}$$

*Define the best effective quality by*

$$Q \equiv \sup_k\{u_k\xi_k\}, \tag{49}$$

*with $Q = 0$ when the point process is empty. Then, for every $q > 0$,*

$$\Pr(Q \le q) = \exp\left\{-\Gamma\left(1-\frac{\theta}{\vartheta}\right)\Xi q^{-\theta}\right\}. \tag{50}$$

*Thus $Q$ is Fréchet with shape $\theta$ and exponent scale $\Gamma(1-\theta/\vartheta)\Xi$.*

**Proof.** The proof has three steps.

*Step 1: compute the applicability moment.* Let $T \equiv \xi^{-\vartheta}$. For every $t \geq 0$,

$$\begin{aligned}\Pr(T > t) &= \Pr(\xi^{-\vartheta} > t) \\ &= \Pr(\xi < t^{-1/\vartheta}) \\ &= \exp\left[-(t^{-1/\vartheta})^{-\vartheta}\right] \\ &= e^{-t}.\end{aligned}$$

At $t = 0$, both sides equal one by continuity. Hence $T \sim$ Exponential(1). Set $r \equiv \theta/\vartheta$. The restriction $\vartheta > \theta$ implies $r \in (0, 1)$, so the following integral is finite:

$$\begin{aligned}\mathbb{E}[\xi^{\theta}] &= \mathbb{E}[(\xi^{-\vartheta})^{-\theta/\vartheta}] \\ &= \mathbb{E}[T^{-r}] \\ &= \int_0^{\infty} t^{-r} e^{-t}\, dt \\ &= \Gamma(1 - r) \\ &= \Gamma\left(1 - \frac{\theta}{\vartheta}\right). \end{aligned} \tag{51}$$

The third equality uses the density $e^{-t}$ of an exponential random variable, and the fourth is the definition of the Gamma function.

*Step 2: compute exactly the mean number of effective ideas above $q$.* For $q > 0$, let

$$N_q \equiv \#\{k : u_k \xi_k > q\}.$$

The marking theorem for Poisson point processes implies that $N_q$ is Poisson. Its mean equals the integral of the indicator $\mathbf{1}\{u\xi > q\}$ against the product of the intrinsic-quality intensity and the mark distribution. Since the integrand is nonnegative, Tonelli's theorem permits changing the order of integration:

$$\begin{aligned}\mathbb{E}[N_q] &= \int_0^{\infty} \int_0^{\infty} \mathbf{1}\{u\xi > q\} \theta \Xi u^{-\theta-1}\, du\, dF_{\xi}(\xi) \\ &= \int_0^{\infty} \left[\int_{q/\xi}^{\infty} \theta \Xi u^{-\theta-1}\, du\right] dF_{\xi}(\xi) \\ &= \int_0^{\infty} \Xi (q/\xi)^{-\theta}\, dF_{\xi}(\xi) \\ &= \Xi q^{-\theta} \mathbb{E}[\xi^{\theta}] \\ &= \Gamma\left(1 - \frac{\theta}{\vartheta}\right) \Xi q^{-\theta}. \end{aligned} \tag{52}$$

The third equality uses

$$\int_{q/\xi}^{\infty} \theta u^{-\theta-1}\, du = (q/\xi)^{-\theta},$$

and the final equality substitutes (51).

*Step 3: apply the Poisson void probability.* The event $\{Q \leq q\}$ is exactly the event that no effective idea exceeds $q$, namely $\{N_q = 0\}$. Because $N_q$ is Poisson with mean given by (52),

$$\begin{aligned}\Pr(Q \leq q) &= \Pr(N_q = 0) \\ &= \exp\{-\mathbb{E}[N_q]\} \\ &= \exp\left\{-\Gamma\left(1 - \frac{\theta}{\vartheta}\right)\Xi q^{-\theta}\right\}.\end{aligned}$$

This is equation (50). If $\Xi = 0$, then $N_q = 0$ almost surely, $Q = 0$, and the same formula gives $\Pr(Q \leq q) = 1$ for every $q > 0$. Thus the result also covers the empty-cohort boundary case. ■

**Innovation and imitation draws.** We now apply Lemma 4 to the two new idea sets that arrive at date $t$.

Innovation in $n$ at date $t$ has intrinsic-quality intensity

$$\theta R_{n,t} u^{-\theta-1}\, du. \tag{53}$$

Setting $\Xi = R_{n,t}$ in (50) gives, for every $q > 0$,

$$\Pr(q_{n,t}^{\text{innov}} \leq q) = \exp\{-A_{n,t}^{\text{innov}} q^{-\theta}\}, \tag{54}$$

$$A_{n,t}^{\text{innov}} \equiv \Gamma\left(1 - \frac{\theta}{\vartheta}\right) R_{n,t}. \tag{55}$$

For imitation, define the newly learned share of cohort $(i, s)$ between the beginnings of dates $t-1$ and $t$ by

$$\Delta\sigma_{in}^{s,t} \equiv \sigma_{in}^{s,t} - \sigma_{in}^{s,t-1}. \tag{56}$$

Substituting the exact diffusion law (41) into this definition yields

$$\Delta\sigma_{in}^{s,t} = \left(1 - \sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} s_{n,t}. \tag{57}$$

The intrinsic-quality intensity of all ideas newly learned in $n$ at date $t$ is therefore

$$\begin{aligned}\Xi_{n,t}^{\text{im}} &\equiv \sum_{i \neq n} \sum_{s<t} R_{i,s} \Delta\sigma_{in}^{s,t} \\ &= \sum_{i \neq n} \sum_{s<t} R_{i,s} \left(1 - \sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} s_{n,t} \\ &= s_{n,t} \sum_{i \neq n} \sum_{s<t} \left(1 - \sigma_{in}^{s,t-1}\right) G_{in}^{s,t-1} R_{i,s} \\ &= s_{n,t} A_{n,t}^{\text{pool}}.\end{aligned} \tag{58}$$

The second equality substitutes (57); the third factors out the destination-date scalar $s_{n,t}$; and the last uses the exact definition (40). Applying Lemma 4 with $\Xi = \Xi_{n,t}^{\text{im}}$ gives

$$\Pr(q_{n,t}^{\text{im}} \leq q) = \exp\{-A_{n,t}^{\text{im}} q^{-\theta}\}, \tag{59}$$

$$\begin{aligned}A_{n,t}^{\text{im}} &= \Gamma\left(1 - \frac{\theta}{\vartheta}\right) \Xi_{n,t}^{\text{im}} \\ &= \Gamma\left(1 - \frac{\theta}{\vartheta}\right) s_{n,t} A_{n,t}^{\text{pool}}.\end{aligned} \tag{60}$$

Equation (60) is identical to (46); it is restated here only to make the frontier recursion self-contained.

**Why the three frontier components are independent.** Date-$t$ innovations are a new Poisson cohort and are independent of all cohorts born before $t$. Within every older cohort, each idea receives one of three mutually exclusive labels at date $t$: known before $t$, newly learned at $t$, or still unknown after learning at $t$. The model assigns these labels by independent Poisson thinning. To verify the resulting independence, fix any effective-quality set $B$ with finite intensity, and let $N_j(B)$ denote the number of ideas in category $j \in \{0,1,2\}$ whose effective quality lies in $B$. Let $\Lambda_j(B)$ be the corresponding thinned intensity. The joint probability generating function supplied by Poisson thinning is

$$\begin{aligned}\mathbb{E}\Big[z_0^{N_0(B)} z_1^{N_1(B)} z_2^{N_2(B)}\Big] &= \exp\left\{\sum_{j=0}^{2} \Lambda_j(B)(z_j - 1)\right\} \\ &= \prod_{j=0}^{2} \exp\{\Lambda_j(B)(z_j - 1)\}\,. \end{aligned} \tag{61}$$

The final expression is the product of the three marginal probability generating functions. Hence the three thinned counts are mutually independent. Taking $B = (q, \infty)$ for every $q > 0$ establishes the conditional independence of the incumbent maximum, the newly learned imitation maximum, and, using independence of the new cohort, the innovation maximum.

**Proposition 5 (Exact frontier-scale recursion)** *Under the independent Poisson marking and thinning described above, suppose the incumbent frontier satisfies*

$$\Pr(q_{n,t-1} \le q) = \exp\{-A_{n,t-1} q^{-\theta}\}, \qquad q > 0, \tag{62}$$

*then*

$$q_{n,t} = \max\left\{q_{n,t-1}, q_{n,t}^{\text{innov}}, q_{n,t}^{\text{im}}\right\} \tag{63}$$

*is Fréchet with shape $\theta$ and exponent scale*

$$A_{n,t} = A_{n,t-1} + A_{n,t}^{\text{innov}} + A_{n,t}^{\text{im}}. \tag{64}$$

*Equivalently,*

$$\boxed{A_{n,t} = A_{n,t-1} + \Gamma\left(1 - \frac{\theta}{\vartheta}\right) R_{n,t} + A_{n,t}^{\text{im}}.} \tag{65}$$

**Proof.** For any $q > 0$, the definition (63) implies the event identity

$$\{q_{n,t} \le q\} = \{q_{n,t-1} \le q\} \cap \{q_{n,t}^{\text{innov}} \le q\} \cap \{q_{n,t}^{\text{im}} \le q\}.$$

Conditional independence of the three subprocesses therefore gives the exact product of their cumulative distribution functions:

$$\begin{aligned}\Pr(q_{n,t} \le q) &= \Pr(q_{n,t-1} \le q)\Pr(q_{n,t}^{\text{innov}} \le q)\Pr(q_{n,t}^{\text{im}} \le q) \\ &= \exp\{-A_{n,t-1} q^{-\theta}\}\exp\{-A_{n,t}^{\text{innov}} q^{-\theta}\}\exp\{-A_{n,t}^{\text{im}} q^{-\theta}\} \\ &= \exp\Big\{-\left(A_{n,t-1} + A_{n,t}^{\text{innov}} + A_{n,t}^{\text{im}}\right) q^{-\theta}\Big\}. \end{aligned} \tag{66}$$

The second equality substitutes (62), (54), and (59). The third equality uses $e^{-a}e^{-b}e^{-c} = e^{-(a+b+c)}$, an exact identity. Comparing (66) with the Fréchet cumulative distribution function $\exp(-A_{n,t}q^{-\theta})$ proves (64). Finally, substituting (55) into (64) gives (65), which is equation (21) in the main text. ■

The recursion is exact because maxima of independent Fréchet draws with a common shape parameter preserve that shape and add their exponent scales. No first-order expansion, log-linearization, or moment approximation is used.

# G Model Calibration and Counterfactual Algorithms

This appendix states the calibration moments, the baseline-equilibrium inversion, and the counterfactual solution procedure used to produce the model results in Sections 4–5. The calibration takes the reduced-form coefficients reported in the paper as estimated moments; it does not re-estimate those regressions jointly with the structural model.

## G.1 Province Innovation and Imitation Levels

The objects used to identify policy responses are distinct from the level moments used to invert the baseline BGP. Let $f$ index firms, let $\pi_{fct}$ denote the technology-innovation index defined in Appendix D, and let $\widehat{\omega}_{fct}$ denote estimated log TFP. The productivity projection gives $\widehat{\beta} = 0.07003598$. At the firm-year level, define the positive innovation and non-own-innovation components by

$$I^0_{fct} = \exp(\widehat{\beta}\pi_{fct}), \qquad M^0_{fct} = \exp(\widehat{\omega}_{fct} - \widehat{\beta}\pi_{fct}). \tag{67}$$

The second component retains the entire part of TFP not attributed to own patenting. In particular, it retains persistent firm and year components; it is not the within-firm residual used as the imitation outcome in the policy regressions.

Let $e_{fct} > 0$ be firm employment, and let $\mathcal{F}_{ct}$ be the common sample with observed $\widehat{\omega}_{fct}$, $\pi_{fct}$, and employment. City-year level moments are constructed by exponentiating first and then taking employment-weighted means:

$$I^0_{ct} = \frac{\sum_{f\in\mathcal{F}_{ct}} e_{fct} I^0_{fct}}{\sum_{f\in\mathcal{F}_{ct}} e_{fct}}, \qquad M^0_{ct} = \frac{\sum_{f\in\mathcal{F}_{ct}} e_{fct} M^0_{fct}}{\sum_{f\in\mathcal{F}_{ct}} e_{fct}}. \tag{68}$$

The baseline is centered on 2010. Because the TFP measure is unavailable in 2010, the implementation averages each city's available level moments over 2009 and 2011. More generally, define

$$\mathcal{T}_c = \left\{ t \in \{2009, 2010, 2011\} : I^0_{ct} \text{ and } M^0_{ct} \text{ are observed} \right\}.$$

$$\overline{I}^0_c = \frac{1}{|\mathcal{T}_c|} \sum_{t\in\mathcal{T}_c} I^0_{ct}, \qquad \overline{M}^0_c = \frac{1}{|\mathcal{T}_c|} \sum_{t\in\mathcal{T}_c} M^0_{ct}. \tag{69}$$

Let $E_{c,2010}$ be total city employment from the 2010 city statistical yearbooks. Define the within-province employment weight as

$$\omega_{cn} = \frac{E_{c,2010}}{\sum_{h\in n} E_{h,2010}}.$$

The province moments are

$$I_n^0 = \sum_{c \in n} \omega_{cn} \overline{I}_c^0, \qquad M_n^0 = \sum_{c \in n} \omega_{cn} \overline{M}_c^0. \tag{70}$$

Equations (67)–(70) describe the exact aggregation used to construct the Chinese province-level baseline moments.

## G.2 Reduced-Form Moments Used by the Model

The baseline calibration uses Table 3 and Table 1. The innovation-choice response is kept in the units of the reported technology-innovation index, because the research-allocation equation is a choice equation rather than a TFP accounting identity. The policy moments are the reported full-post-window level effects, not coefficients re-estimated on a five-year post-treatment window; the latter are retained as a robustness specification. The moments are

$$\phi_R = 0.049, \qquad \phi_M = -0.074, \qquad \rho_M^K = 0.229,$$

and the policy-by-access interactions are

$$\phi_R^K = -0.021, \qquad \phi_M^K = -0.084.$$

The common slope ratio at the sample mean is therefore

$$c_0 = -\frac{\phi_M}{\phi_R} = 1.510.$$

Province heterogeneity is identified from the same Table 3 interactions. To preserve the centering of that regression, let $\mathcal{S}_0$ denote the firm-year panel after applying its treatment-timing restrictions, and let $\mathcal{S} \subseteq \mathcal{S}_0$ denote observations with nonmissing TFP, treatment, access, and controls. If $K_{ct}$ is accessible knowledge, define

$$\overline{K} = \frac{1}{|\mathcal{S}_0|} \sum_{(f,c,t) \in \mathcal{S}_0} K_{ct}, \qquad \widetilde{K}_n = \frac{1}{|\mathcal{S}_n|} \sum_{(f,c,t) \in \mathcal{S}_n} K_{ct} - \overline{K}, \tag{71}$$

where $\mathcal{S}_n$ contains the observations in $\mathcal{S}$ located in province $n$. Thus the province measure retains the firm-year weighting and the evaluation point of the reported regression; it is not re-centered across the thirty provinces. For provinces absent from $\mathcal{S}$ because all required controls are missing, the corresponding city-year mean of $K_{ct}$ over 2000–2013 is used and is centered by the same $\overline{K}$.

Because the city-level estimates do not identify responses arbitrarily far outside their empirical support, define

$$\widetilde{K}_n^\star = \max\{\widetilde{K}_n, -1\}.$$

The heterogeneous parameters follow the log schedules

$$c_n = c_0 \exp(\gamma_c \widetilde{K}_n^\star), \qquad \chi_n = \chi_0 \exp(\gamma_\chi \widetilde{K}_n^\star),$$

where

$$\gamma_c = \frac{-\phi_M^K}{-\phi_M} + \frac{-\phi_R^K}{\phi_R}.$$

The restriction affects only provinces with $\widetilde{K}_n < -1$ and prevents the province aggregation from extrapolating the estimated city-level interaction beyond the lower tail of the estimation sample. The specification without this support restriction is reported as a robustness exercise. Given $\nu = 0.76$, $\eta$ is pinned by the access-only imitation elasticity:

$$\rho_M^K = \frac{(1-\eta)\{c_0 + (1-\nu)\eta\}}{(1-\nu\eta)c_0 + (1-\nu)\eta}.$$

The policy pass-through at the sample mean is

$$\chi_0 = \frac{\phi_R\{(1-\nu\eta)c_0 + (1-\nu)\eta\}}{\nu\eta},$$

and the access gradient in policy pass-through is

$$s_0 = \frac{(1-\nu\eta)c_0}{(1-\nu\eta)c_0 + (1-\nu)\eta}, \qquad \gamma_\chi = \frac{\phi_R^K}{\phi_R} + s_0\gamma_c. \tag{72}$$

These equations yield the parameter values reported in Table 8.

### G.3 Common-Unit Normalization

The province levels in equation (70) are positive, but their relative scale is not identified by the firm-level decomposition. This is especially important because $\phi_R$ is kept in the units of the raw technology-innovation choice index, whereas $M_n^0$ is constructed from log TFP. The calibration therefore applies one common scalar to imitation,

$$M_n(\lambda_M) = \lambda_M M_n^0, \qquad \Omega_n^0(\lambda_M) = c_n \frac{M_n(\lambda_M)}{I_n^0} = \lambda_M c_n \frac{M_n^0}{I_n^0}.$$

The scalar $\lambda_M$ is identified from the Table 3 TFP policy-by-access moments,

$$b_0 = -0.058, \qquad b_K = -0.097.$$

For a trial $\lambda_M$, the model computes the local marginal frontier effect of a policy wedge using the research-allocation FOC:

$$b_n^A(\lambda_M) = \chi_n \frac{\nu\eta I_n^0 M_n(\lambda_M)\{1-\Omega_n^0(\lambda_M)\}}{\{(1-\nu\eta)\Omega_n^0(\lambda_M)I_n^0 + (1-\nu)\eta M_n(\lambda_M)\}\{I_n^0 + M_n(\lambda_M)\}}.$$

The projection uses all thirty Chinese provinces with equal weight and the untruncated $\widetilde{K}_n$ from equation (71). Let $N_C = 30$, $\overline{\widetilde{K}} = N_C^{-1}\sum_n \widetilde{K}_n$, and $\bar{b}^A(\lambda_M) = N_C^{-1}\sum_n b_n^A(\lambda_M)$. The projection coefficients are computed exactly as

$$m_K(\lambda_M) = \frac{\sum_n (\widetilde{K}_n - \overline{\widetilde{K}}) b_n^A(\lambda_M)}{\sum_n (\widetilde{K}_n - \overline{\widetilde{K}})^2}, \tag{73}$$

$$m_0(\lambda_M) = \bar{b}^A(\lambda_M) - m_K(\lambda_M)\overline{\widetilde{K}}. \tag{74}$$

The scalar $\lambda_M$ solves the zero-crossing condition implied by the reported TFP regression,

$$m_0(\lambda_M)b_K - b_0 m_K(\lambda_M) = 0.$$

This pins the location where the model's marginal policy effect changes sign to the same location implied by the empirical TFP slope and interaction. The preferred solution is $\lambda_M = 0.01061594$, which gives $\Omega_n^0 \in [0.246, 6.588]$ across Chinese provinces.

## G.4 Baseline Equilibrium Inversion

The baseline economy has thirty Chinese provinces plus rest of world. The calibration uses province trade shares, production labor, researcher endowments, income, trade deficits, and the innovation and imitation moments constructed above. Define $\Gamma_q = \Gamma(1 - \theta/\vartheta)$ and set $I_n = I_n^0$ and $M_n = M_n(\lambda_M)$ for Chinese provinces.

**Research allocation and flow productivities.** Because $\Omega_n^0 I_n = c_n M_n$, the research-allocation FOC and $L_n^I + L_n^M = \bar{L}_n^R$ imply

$$\frac{L_n^M}{L_n^I} = \frac{\eta}{c_n}, \qquad L_n^I = \frac{\bar{L}_n^R}{1 + \eta/c_n}, \qquad L_n^M = \frac{(\eta/c_n)\bar{L}_n^R}{1 + \eta/c_n}. \tag{75}$$

The innovation flow and innovation productivity are then

$$R_n = \frac{I_n}{\Gamma_q}, \qquad a_n^I = \frac{I_n/\Gamma_q}{(L_n^I)^\nu}. \tag{76}$$

**Imitable-pool fixed point.** For Chinese provinces, the fixed point solved in the calibration is

$$s_n = \min\left\{\frac{M_n}{\Gamma_q A_n^{pool}}, d\right\}, \qquad \sigma = \mathcal{S}(\lambda^{data}, s), \qquad A^{pool} = \mathcal{P}(\sigma, R), \tag{77}$$

where $\mathcal{S}$ is the diffusion recursion and $\mathcal{P}$ is the cohort sum defined in Appendix F.1. Starting from a strictly positive pool, iteration $k$ computes $s^{(k)}$ and $\sigma^{(k)}$ from the first two equations, forms $\widetilde{A}^{pool,(k+1)} = \mathcal{P}(\sigma^{(k)}, R)$, and updates

$$A^{pool,(k+1)} = 0.5A^{pool,(k)} + 0.5\widetilde{A}^{pool,(k+1)}. \tag{78}$$

Iteration stops when the maximum relative pool change is below $10^{-9}$. The upper bound in equation (77) does not bind in the baseline calibration. Imitation productivity is therefore recovered as

$$a_n^M = \left[\frac{M_n/\Gamma_q}{d(A_n^{pool})^{1-\eta}(L_n^M)^{\nu\eta}}\right]^{1/\eta}. \tag{79}$$

The frontier scale follows exactly from

$$A_n = \frac{I_n + M_n}{1 - g^{-\nu}}. \tag{80}$$

Rest of world is treated differently because its imitation flow is not observed. Its innovation flow equals the rest-of-world-to-China resident-patent ratio times aggregate Chinese innovation. At every pool iteration its learning rate is set to the unweighted mean Chinese rate,

$$s_{ROW} = \frac{1}{30}\sum_{n \in China} s_n, \qquad M_{ROW} = \Gamma_q s_{ROW} A_{ROW}^{pool}. \tag{81}$$

Thus rest-of-world imitation is derived from its pool rather than imposed as a fixed imitation-to-innovation ratio.

**Wages, trade costs, migration costs, and amenities.** Let $y_n$ be production-labor income recovered from the trade accounts. Wages are not read as an independent moment; they are constructed as

$$w_n = \frac{y_n/\bar{L}_n^P}{y_{ROW}/\bar{L}_{ROW}^P}, \qquad w_{ROW} = 1, \tag{82}$$

and trade deficits are expressed in the same numeraire. Bilateral trade costs use the Eaton and Kortum (2002) inversion as an initial value and then match the directional trade shares with the correlated-GEV update

$$\tau_{od}^{(k+1)} = \tau_{od}^{(k)} \left( \frac{\lambda_{od}^{model}(\tau^{(k)}; \sigma, R, w)}{\lambda_{od}^{data}} \right)^{0.6/\theta}, \qquad \tau_{dd} = 1, \tag{83}$$

holding $(\sigma, R, w)$ fixed during this inversion. Iteration stops when the maximum absolute trade-share discrepancy is below $10^{-9}$.

Migration costs are recovered symmetrically from the 2010 origin–destination shares,

$$m_{in} = -\frac{\zeta}{2} \log\left( \frac{\mu_{in}^{data} \mu_{ni}^{data}}{\mu_{ii}^{data} \mu_{nn}^{data}} \right), \qquad m_{ii} = 0. \tag{84}$$

Zero or unavailable bilateral cells, including unavailable rest-of-world flows, receive the finite high-cost value $m_{in} = 12$. Given the resulting price index, amenities are inverted so that the stationary migration distribution equals observed production labor. Finally, the recovered primitives are supplied to the full BGP solver. The calibration is accepted only if that forward solution reproduces the targeted trade shares, flows, wages, pools, and population distribution and all equilibrium residuals are numerically negligible.

## G.5 Counterfactual Solver

Every counterfactual keeps primitives $\{a_n^I, a_n^M, \bar{L}_n^R, \tau_{in}, m_{in}, B_n\}$ fixed unless the experiment explicitly changes one of them. The stationary BGP is re-solved after each shock. Large shocks use continuation in the shocked object and warm-start each step from the previous solution.

**Measured subsidy experiments.** The province exposure uses the same post-treatment indicator as the policy regressions. Over 2009–2011, let $D_{ft}$ equal one in the firm's first subsidy year and every subsequent observed year. Let $P_{ft}$ be its total patent applications. Define

$$W_n = \sum_{f \in n} \sum_{t=2009}^{2011} P_{ft}, \qquad Z_n^0 = \frac{\sum_{f \in n} \sum_{t=2009}^{2011} P_{ft} D_{ft}}{W_n}. \tag{85}$$

Thus $Z_n^0$ is the patent-weighted share of province innovation activity exposed to the policy, and $W_n$ is patent mass. It is not observed fiscal expenditure.

The calibrated wedge $\Omega_n^0$ already corresponds to $Z_n^0$. Consequently, every policy counterfactual uses the change in exposure,

$$\Omega_n' = \Omega_n^0 \exp\Big\{ \chi_n (Z_n' - Z_n^0) \Big\}, \tag{86}$$

with rest of world held fixed. The removal experiment sets $Z'_n = 0$. The uniform reallocation holds the aggregate exposure mass $B_Z = \sum_n W_n Z_n^0$ fixed and sets

$$Z'_n = \frac{B_Z}{\sum_j W_j}. \tag{87}$$

The knowledge-based reallocation uses

$$Z'_n = \min\{1, \max\{0, \alpha - \beta_Z \log A_n^{pool}\}\}, \qquad \alpha = \frac{B_Z + \beta_Z \sum_j W_j \log A_j^{pool}}{\sum_j W_j}. \tag{88}$$

For the reported values $\beta_Z \in \{0.1, 0.2\}$, neither bound binds, so $\sum_n W_n Z'_n = B_Z$ holds exactly.

For the marginal-welfare reallocation, the model first raises one province's exposure by $h = 0.2$, re-solves the full BGP, and computes the forward-difference marginal value per unit of exposure mass,

$$g_n = \frac{\{\mathcal{W}(Z^0 + h\mathbf{e}_n) - \mathcal{W}(Z^0)\}/h}{W_n}. \tag{89}$$

The reported reallocation steps move exposure toward provinces with high $g_n$ after subtracting its $W_n$-weighted mean. These are budget-preserving gradient steps, not solutions to the unrestricted planner problem.

**Wedge-dispersion experiments.** The headline experiment changes the baseline wedge directly:

$$\log \Omega'_n = \sum_{j \in China} s_j^R \log \Omega_j^0 \quad \text{for all Chinese provinces } n,$$

holding the research-weighted log mean fixed. The best-common-wedge experiment instead solves the BGP on a grid of common values $\Omega'_n = \omega$ and reports the value with the highest worker welfare.

**Trade experiment.** The trade liberalisation experiment sets $\tau'_{in} = 0.9\tau_{in}$ for off-diagonal bilateral pairs and re-solves the BGP. This affects productivity only through market access and the imitable pool, not through the policy wedge.

### G.6 Reported Outcomes

Aggregate TFP is the production-labor-weighted frontier change,

$$\Delta \ln TFP = \sum_{n \in China} s_n^P \Delta \ln A_n.$$

Worker welfare is the consumption-equivalent change from the dynamic migration value function. Province-level results report changes in $A_n$, $I_n$, $M_n$, research allocation, and welfare. Every reported counterfactual compares two stationary BGP solutions: the calibrated baseline and the equilibrium obtained after changing only the primitive or wedge specified by the experiment. Thus the reported changes include all equilibrium adjustments in wages, prices, trade shares, migration, knowledge diffusion, and the allocation of research labor.

# H Robustness Checks

This appendix reports supplementary robustness exercises for the main empirical results.

**Alternative TFP estimators.** Our main results use the Ackerberg, Caves and Frazer (2015) TFP estimator. Appendix Table A.2 (available upon request) reports equivalent estimates using the Olley and Pakes (1996) estimator. Coefficients are qualitatively and quantitatively similar across both methods.

**Alternative instrument specifications.** For the HSR instrument, we consider (i) the simpler binary variant $Z_{ct}^{\text{simple}} = \text{PlannedNet}_c \times \mathbf{1}[\text{HSR}_{ct} = 1]$, which uses planned-network exposure directly without knowledge-weighting; (ii) a continuous distance-decay variant that replaces the binary $\text{PlannedNet}_c$ indicator with a continuous weight inversely proportional to the distance from city $c$ to the nearest planned-network segment. For the Bartik IV, we consider (i) a broader variant that includes all HS chapters rather than restricting to high-technology sectors; (ii) employment-weighted rather than output-weighted shares. Results are stable across all variants.

**Alternative clustering.** Our main estimates cluster standard errors at the firm level. In Appendix Table A.3 (available upon request) we report estimates clustering at the city level, the city–year level, and using two-way clustering at the city and year levels. Inference is robust throughout.

**Placebo tests.** To assess whether the IV results are driven by pre-existing trends, we implement placebo tests that assign the instruments to firm cohorts that were never treated (i.e., never received a subsidy). The interaction coefficients are statistically indistinguishable from zero in all placebo specifications.

**Controlling for concurrent policies.** We control for the Urban Integration Policy (UIP) and other concurrent regional development programmes that coincide with HSR expansion. Including indicators for UIP adoption and its interaction with our subsidy indicator leaves the main coefficients largely unchanged.

# I Measurement of Breakthrough and Imitative Patents

This appendix describes the construction of the breakthrough (BK) and imitative (Mimic) patent classifications used in Section 3.1, following the methodology of Liu and Wang (2025).

**Data and embedding.** Raw patent records are drawn from the China National Intellectual Property Administration (CNIPA) corpus, processed through a standardized pipeline. Each patent abstract is translated to English (where needed) and embedded into a 1,536-dimensional semantic vector using a pre-trained language model. Cosine similarity between two patents is computed as the dot product of their normalized embedding vectors, measuring the closeness of their technical content.

**Breakthrough (BK) classification.** Following Kelly et al. (2021), the innovation quality score for focal patent $f$ granted in year $y$ is:

$$s_f^{\text{BK}} = \log \frac{\sum_{y+1 \le \text{year}(p) \le y+5} \cos(f,p)}{\sum_{y-5 \le \text{year}(p) \le y-1} \cos(f,p)}, \tag{90}$$

where the sums are over all patents $p$ in the CNIPA corpus. A high $s_f^{\text{BK}}$ indicates that patent $f$ is more similar to future patents than to past patents—it opened a new direction of research rather than elaborating an existing one. Scores are demeaned within technology class and year. Patents in the top decile of the score distribution (China-domestic threshold, within technology class and year) are classified as BK.

**Imitative (Mimic) classification.** The imitation score for focal patent $f$ is:

$$s_f^{\text{MIM}} = \max_{\substack{y-5 \le \text{year}(p) \le \text{year}(f) \\ p \ne f}} \cos(f,p), \tag{91}$$

the maximum cosine similarity between $f$ and any patent granted in the preceding five years. A high $s_f^{\text{MIM}}$ indicates that patent $f$ is very close to recent prior art—it closely copies an existing invention. Patents in the top decile of the similarity distribution (within technology class and year) are classified as Mimic.

**City-year aggregation.** Patent-level BK and Mimic indicators are assigned to prefecture cities using the applicant address. City-year counts $n_{ct}^{\text{BK}}$, $n_{ct}^{\text{Mimic}}$, and $n_{ct}^{\text{total}}$ are summed across all patents attributed to city $c$ in year $t$.

Because both labels select the top decile of their respective distributions, $n_{ct}^{\text{BK}} + n_{ct}^{\text{Mimic}} < n_{ct}^{\text{total}}$ in every city-year: the two categories are the tails of the quality distribution, not a partition of it. The rate outcomes divide by $n_{ct}^{\text{total}}$ (all patents, not the eligible sub-pools), so each rate measures how frequently a city's patents reach one tail, and the two rates do not sum to one. Each is centered near 0.10 by construction, and the identifying variation comes from departures of a city-year from that benchmark. The log ratio $\ln(n_{ct}^{\text{BK}}+1) - \ln(n_{ct}^{\text{Mimic}}+1)$ instead compares the two tails to each other and is unaffected by the size of the untagged middle of the distribution. We report both

because they answer different questions: the rates show whether tail activity of each kind rose or fell, and the ratio shows which tail came to dominate.

# J Iceberg Trade Costs and Accessible Knowledge

This appendix describes the construction of city-level iceberg trade costs and accessible knowledge measures. Both variables play a central role in capturing spatial frictions and external knowledge exposure faced by firms located in different cities.

## J.1 Measuring Iceberg Trade Costs

**Economic motivation.** In models of trade and economic geography, iceberg trade costs capture the loss of goods or efficiency incurred when shipping products across space. Lower iceberg costs facilitate market access, intensify competition, and enhance the diffusion of ideas and technologies. Because direct measures of trade costs are rarely observable at the city level, we proxy iceberg costs using transportation times between cities.

**Baseline measure.** Following Ma and Tang (2024), we measure iceberg trade costs using bilateral road transportation times. Let $T_{i,j}^{\text{road}}$ denote the shortest road travel time between city $i$ and city $j$. We focus on each citys ten geographically nearest cities to capture economically relevant local trade connections.

Formally, iceberg trade costs for city $i$ are defined as:

$$\tau_i = \sum_{j \in \mathcal{N}_{20}(i)} T_{i,j}^{\text{road}}, \tag{92}$$

where $\mathcal{N}_{20}(i)$ denotes the set of the 20 nearest cities to city $i$. A higher value of $\tau_i$ indicates greater spatial frictions and higher iceberg trade costs faced by firms in city $i$.

**Implementation details.** Transportation times are computed using the road network that minimizes travel duration. Restricting attention to the nearest cities ensures that the measure reflects economically meaningful interactions rather than distant, infrequent trade links. In the empirical analysis, $\tau_i$ is standardized to facilitate interpretation and comparability across specifications.

**Robustness and alternative measures.** To assess robustness, we construct alternative iceberg cost measures using railway transportation times. Let $T_{i,j}^{\text{rail}}$ denote the shortest railway travel time between cities $i$ and $j$. We consider two alternatives:

1. A railway-based measure:
$$\tau_i^{\text{rail}} = \sum_{j \in \mathcal{N}_{20}(i)} T_{i,j}^{\text{rail}}; \tag{93}$$

2. A combined measure that averages road and railway transportation times:
$$\tau_i^{\text{avg}} = \sum_{j \in \mathcal{N}_{20}(i)} \frac{1}{2} \left( T_{i,j}^{\text{road}} + T_{i,j}^{\text{rail}} \right). \tag{94}$$

These alternative measures address potential concerns that a single transportation mode may not fully capture trade frictions.

## J.2 Measuring Accessible Knowledge

**Economic motivation.** Knowledge diffusion is inherently spatial: firms learn more readily from nearby cities, especially when road transport is cheap and fast. Our accessible-knowledge measure aggregates the technological strength of surrounding cities, discounted by bilateral road travel cost, into a single city–year index. A key design choice is how to measure each city's "technological

strength." The raw city-average TFP conflates within-industry productivity with the industry composition of the city—a city that happens to concentrate in high-TFP industries will appear productive even if its firms are not particularly efficient within their own industries. To remove this composition bias, we construct the city-average TFP index using firm-level TFP *z-scored within narrow industry×year cells* before aggregating to the city level.

**Step 1: Industry-adjusted firm TFP (z-score).** Let $\omega_{ict}$ denote the ACF TFP estimate for firm $i$ in city $c$ in year $t$, estimated separately by two-digit CIC industry. We further demean within *small-category* (four-digit CIC) industry×year cells. Specifically, for each small-category industry $s$ and year $t$, let $\mu_{st}$ and $\sigma_{st}$ denote the cross-firm mean and standard deviation of $\omega_{ict}$ within that cell. The industry-adjusted TFP z-score is:

$$z_{ict} = \frac{\omega_{ict} - \mu_{st}}{\sigma_{st}}, \tag{95}$$

with cells containing fewer than five observations excluded to ensure stable estimates. This transformation preserves within-industry cross-firm productivity differences while removing variation attributable to industry composition.

**Step 2: City-average industry-adjusted TFP.** The city-year TFP index $\bar{\omega}_{ct}$ is then the simple mean of the z-scores across all firms in city $c$ in year $t$:

$$\bar{\omega}_{ct} = \frac{1}{|\mathcal{F}_{ct}|} \sum_{i \in \mathcal{F}_{ct}} z_{ict}, \tag{96}$$

where $\mathcal{F}_{ct}$ is the set of firms in city $c$ in year $t$. Because each $z_{ict}$ is standardized within its industry×year cell, $\bar{\omega}_{ct}$ reflects the average within-industry productivity *standing* of firms located in city $c$—not the productivity premium of being in a high-TFP industry.

**Step 3: Accessible knowledge index.** The accessible knowledge index for city $c$ in year $t$ aggregates the industry-adjusted TFP of the thirty lowest-cost outside cities, discounted by road travel cost:

$$A_{ct} = \ln\left(1 + \sum_{d \in \mathcal{N}_{30}(c,t)} \frac{\exp(\bar{\omega}_{dt})}{r_{cdt}}\right), \tag{97}$$

where $r_{cdt}$ is the bilateral road travel cost (in hours) from city $c$ to city $d$ in year $t$, and $\mathcal{N}_{30}(c,t)$ denotes the set of the 30 outside cities with the lowest road cost from $c$ in year $t$. The exponential transformation maps $\bar{\omega}_{dt}$ back into a productivity-level index; dividing by $r_{cdt}$ discounts more distant cities; the logarithm compresses the right tail. Using 30 neighbors (rather than a smaller number) ensures that the index captures the broad regional knowledge pool available to a city rather than only its immediate neighbors, while the cost-based neighbor set ensures that the 30 cities are those most economically relevant to city $c$.

**Comparison with alternative constructions.** Two aspects of our construction distinguish it from common alternatives. *First*, the z-scoring step is critical. Using raw average TFP $\bar{\omega}_{ct}^{\text{raw}}$ instead of the industry-adjusted index $\bar{\omega}_{ct}$ conflates knowledge access with industrial composition:

coastal cities with large high-tech sectors appear to offer richer knowledge simply because their industry mix is favorable, even if within-industry efficiency is no higher than inland cities. The z-score strips this out, so $A_{ct}$ captures cross-city differences in within-industry productive knowledge rather than inter-industry compositional differences. *Second*, limiting the neighbor set to the 30 lowest-cost (rather than 30 nearest great-circle) cities allows the accessible knowledge pool to update as transport infrastructure improves—HSR opening, for instance, can bring a previously distant hub into a city's effective neighborhood. This feature links $A_{ct}$ directly to the HSR instrument constructed in Appendix E.

**Empirical properties.** The resulting index exhibits a pronounced East–West gradient: coastal cities in Guangdong, Zhejiang, and Jiangsu score in the top quartile throughout the sample, reflecting both their own high within-industry TFP and their dense road connections to other productive cities. Northwestern and Tibetan cities consistently score in the bottom quartile. The index trends upward over 2000–2013 as road infrastructure improves and firm productivity grows, with the steepest gains in cities adjacent to newly opened expressway segments—consistent with the instrument logic in Appendix E. Figure 14 maps the index in 2013.

# K Patent-Based Knowledge Diffusion

This appendix describes the construction of our patent-based measures of technological similarity and imitation. The objective is to quantify both the technological proximity across cities and the realized knowledge flows embodied in patent citations, and to aggregate these measures to the city–year level in a manner consistent with the empirical analysis.

## K.1 Patent-Level Technological Similarity

Each patent $p$ is represented by a fixed-dimensional text embedding vector $\mathbf{e}_p \in \mathbb{R}^d$. All embeddings are $\ell_2$-normalized at the patent level. Technological similarity between two patents $p$ and $q$ is measured using cosine similarity,

$$s_{pq} = \frac{\mathbf{e}_p^\top \mathbf{e}_q}{\|\mathbf{e}_p\| \, \|\mathbf{e}_q\|}. \tag{98}$$

To isolate economically meaningful technological proximity, we introduce a similarity threshold based on the upper tail of the empirical similarity distribution. We randomly draw a large subset of Chinese patent applications and form a large random sample of unordered patent pairs by independently drawing two distinct patents from this subset. Computing cosine similarity for each pair yields an empirical distribution of patent-pair similarities that is not conditioned on city or year.

We define the similarity threshold $\tau$ as the 90th percentile of this distribution. In the data, the estimated threshold is

$$\tau = 0.3876908.$$

All similarity-based measures use this single threshold throughout the analysis.

## K.2 City–City Technological Similarity Measures

Patents are aggregated at the city–year level. Let $\mathcal{P}_{c,t}$ denote the set of patents applied for in city $c$ in year $t$. For any ordered pair of city–years $(c_1, t_1)$ and $(c_2, t_2)$, we consider all patent pairs $(p, q)$ with $p \in \mathcal{P}_{c_1,t_1}$ and $q \in \mathcal{P}_{c_2,t_2}$. When $t_1 = t_2$, patent pairs originating from the same city are excluded to avoid mechanically inflated similarity.

We construct three complementary measures of technological proximity across cities.

**Similar Patent: Count.** The extensive margin of technological overlap is measured as

$$\text{SPCount}_{c_1,t_1,c_2,t_2} = \sum_{p \in \mathcal{P}_{c_1,t_1}} \sum_{q \in \mathcal{P}_{c_2,t_2}} \mathbf{1}\{s_{pq} \geq \tau\}. \tag{99}$$

**Similar Patent: Percentage.** To scale the count by the number of potential patent pairs, we define

$$\text{SPShare}_{c_1,t_1,c_2,t_2} = \frac{\text{SPCount}_{c_1,t_1,c_2,t_2}}{|\mathcal{P}_{c_1,t_1}| \cdot |\mathcal{P}_{c_2,t_2}|}. \tag{100}$$

**Conditional Similarity Enhancement.** To capture the intensity of technological proximity conditional on exceeding the similarity threshold, we define

$$\text{CSE}_{c_1,t_1,c_2,t_2} = \frac{\sum_{p \in \mathcal{P}_{c_1,t_1}} \sum_{q \in \mathcal{P}_{c_2,t_2}} s_{pq}\, \mathbf{1}\{s_{pq} \geq \tau\}}{\text{SPCount}_{c_1,t_1,c_2,t_2}}. \tag{101}$$

In the empirical analysis, we use the logarithm of this measure to reduce skewness.

## K.3 Citation-Based Measures of Knowledge Diffusion

Technological similarity captures the feasibility of learning, but does not by itself reflect realized knowledge flows. We therefore construct knowledge diffusion measures based on inter-city patent citations.

Let $\mathcal{C}_{c_1,t_1,c_2,t_2}$ denote the set of citations from patents applied for in city $c_1$ in year $t_1$ to patents applied for in city $c_2$ in year $t_2$. Citations are indexed by origin city, destination city, application year, and citation lag. We aggregate citations at the city–year–lag level, allowing for flexible citation horizons.

Because citations require both citing and cited patents, citation-based imitation is inherently conditional on positive patenting activity in both city–years. As a result, the resulting citation objects do not form a balanced panel. This feature mirrors the construction of the similarity measures, which are defined only when patent applications exist in both cities.

## K.4 City-Level Aggregation of Residual-Based Imitation

Building on the residual-based firm-level imitation measure, we aggregate imitation to the city–year level.

Let $\mathcal{F}_{c,t}$ denote the set of firms located in city $c$ in year $t$. City-level imitation intensity is defined as the sum of firm-level imitation components within a city:

$$\text{Imitation}_{c,t} = \sum_{i \in \mathcal{F}_{c,t}} \text{Immi}_{it}, \tag{102}$$

where $\text{Immi}_{it} \equiv \widehat{\varepsilon}_{it}$ is the residual from the firm-level productivity regression.

## References


**Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti.** 2006. "Distance to Frontier, Selection, and Economic Growth." *Journal of the European Economic Association*, 4(1): 37–74. 5

**Ackerberg, Daniel A., Kevin Caves, and Garth Frazer.** 2015. "Identification properties of recent production function estimators." *Econometrica*, 83(6): 2411–2451. 7, 42, 44, 49, 50, 51, 65

**Arrow, Kenneth J.** 1962. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*. 609–626. Princeton, NJ:Princeton University Press. 2

**Benhabib, Jess, Jesse Perla, and Christopher Tonetti.** 2014. "Catch-up and Fall-back through Innovation and Imitation." *Journal of Economic Growth*, 19(1): 1–35. 5

**Benhabib, Jess, Jesse Perla, and Christopher Tonetti.** 2021. "Reconciling Models of Diffusion and Innovation: A Theory of the Productivity Distribution and Technology Frontier." *Econometrica*, 89(5): 2261–2301. 5

**Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb.** 2020. "Are ideas getting harder to find?" *American Economic Review*, 110(4): 1104–1144. 26, 28

**Bloom, Nicholas, Mark Schankerman, and John Van Reenen.** 2013. "Identifying technology spillovers and product market rivalry." *Econometrica*, 81(4): 1347–1393. 2, 5

**Buera, Francisco J., and Ezra Oberfield.** 2020. "The Global Diffusion of Ideas." *Econometrica*, 88(1): 83–114. 5

**Busso, Matías, Jesse Gregory, and Patrick Kline.** 2013. "Assessing the incidence and efficiency of a prominent place-based policy." *American Economic Review*, 103(2): 897–947. 5

**Cai, Sheng, Ningyuan Jia, Weizhe Wang, and Wei Xiang.** 2025. "The Innovation–Imitation Tradeoff: Micro Evidence, Macro Implications." Working Paper. 3, 8, 16

**Cohen, Wesley M., and Daniel A. Levinthal.** 1989. "Innovation and Learning: The Two Faces of R&D." *The Economic Journal*, 99(397): 569–596. 2

**Eaton, Jonathan, and Samuel Kortum.** 2002. "Technology, geography, and trade." *Econometrica*, 70(5): 1741–1779. 4, 23, 24, 25, 63

**Fajgelbaum, Pablo D., and Cecile Gaubert.** 2020. "Optimal spatial policies, geography, and sorting." *Quarterly Journal of Economics*, 135(2): 959–1036. 5

**Gaubert, Cecile.** 2018. "Firm sorting and agglomeration." *American Economic Review*, 108(11): 3117–3153. 5

**Griffith, Rachel, Stephen Redding, and John Van Reenen.** 2004. "Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries." *Review of Economics and Statistics*, 86(4): 883–895. 2

**Howell, Sabrina T.** 2017. "Financing innovation: Evidence from R&D grants." *American Economic Review*, 107(4): 1136–1164. 5

**Howitt, Peter, and David Mayer-Foulkes.** 2005. "R&D, Implementation, and Stagnation: A Schumpeterian Theory of Convergence Clubs." *Journal of Money, Credit and Banking*, 37(1): 147–177. 5

**Hsieh, Chang-Tai, and Peter J. Klenow.** 2009. "Misallocation and manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–1448. 5

**Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow.** 2019. "The allocation of talent and U.S. economic growth." *Econometrica*, 87(5): 1439–1474. 5

**Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson.** 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *Quarterly Journal of Economics*, 108(3): 577–598. 5

**Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy.** 2021. "Measuring Technological Innovation over the Long Run." *American Economic Review: Insights*, 3(3): 303–320. 13, 14, 66

**Kline, Patrick, and Enrico Moretti.** 2014. "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority." *Quarterly Journal of Economics*, 129(1): 275–331. 5

**König, Michael D., Jan Lorenz, and Fabrizio Zilibotti.** 2016. "Innovation vs. Imitation and the Evolution of Productivity Distributions." *Theoretical Economics*, 11(3): 1053–1102. 5

**König, Michael, Kjetil Storesletten, Zheng Song, and Fabrizio Zilibotti.** 2022. "From Imitation to Innovation: Where Is All That Chinese R&D Going?" *Econometrica*, 90(4): 1615–1654. 5

**Lind, Nelson, and Natalia Ramondo.** 2023. "Trade with correlation." *American Economic Review*, 113(2): 317–353. 4, 25, 28

**Liu, Ernest.** 2019. "Industrial policies in production networks." *Quarterly Journal of Economics*, 134(4): 1883–1948. 5

**Liu, Huben, and Weizhe Wang.** 2025. "Global Innovation Dynamics over the Past Three Decades." Working Paper. 2, 13, 66

**Lucas, Robert E., Jr., and Benjamin Moll.** 2014. "Knowledge Growth and the Allocation of Time." *Journal of Political Economy*, 122(1): 1–51. 5

**Ma, Lin, and Yang Tang.** 2024. "The distributional impacts of transportation networks in China." *Journal of International Economics*, 148. Article in Press. 68

**Moretti, Enrico.** 2004. "Workers' education, spillovers, and productivity: Evidence from plant-level production functions." *American Economic Review*, 94(3): 656–690. 5

**Nelson, Richard R.** 1959. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy*, 67(3): 297–306. 2

**Olley, Steven, and Ariel Pakes.** 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64: 1263–1297. 49, 65

**Perla, Jesse, and Christopher Tonetti.** 2014. "Equilibrium Imitation and Growth." *Journal of Political Economy*, 122(1): 52–76. 5

**Perla, Jesse, Christopher Tonetti, and Michael E. Waugh.** 2021. "Equilibrium Technology Diffusion, Trade, and Growth." *American Economic Review*, 111(1): 73–128. 5

**Sampson, Thomas.** 2016. "Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth." *Quarterly Journal of Economics*, 131(1): 315–380. 5

**Tombe, Trevor, and Xiaodong Zhu.** 2019. "Trade, Migration, and Productivity: A Quantitative Analysis of China." *American Economic Review*, 109(5): 1843–1872. 26, 28

**Xiang, Wei.** 2025. "Clean Growth and Environmental Policies in the Global Economy." *Available at SSRN 4619717*. 26, 28