How Knowledge Diffusion Shapes Innovation Composition: Evidence from China

Yousha Liang Kang Shi Junjie Tang Hanyi Tao Juanyi Xu*

July 2, 2025

Abstract

This paper explores how knowledge diffusion through high-speed rail (HSR) network shapes the composition of local innovations. Leveraging a large-scale textual analysis of Chinese patent data and quasi-experimental variation from China's HSR expansion, we find that enhanced inter-city HSR connectivity disproportionately fosters incremental innovation over breakthrough innovation. This is because incremental innovation is intensive in existing knowledge stocks. By accelerating technology spillovers, HSR increases innovators' access to distant knowledge and reduces the relative cost of incremental innovations compared to breakthrough projects. Moreover, accelerated knowledge flows affect the direction of local innovation, resulting in greater similarity in innovations between HSR-connected cities.

JEL Code: F13, F14, O14, O31.

Keywords: Innovation, Incremental innovation, Technology spillovers, Textual analysis, High-speed rail

*Liang: School of Economics, Zhejiang University, Hangzhou, China. Email: youshaliang@zju.edu.cn. Shi: Department of Economics, Chinese University of Hong Kong, Shatin, New Territories, Hong Kong. Email: kangshi@cuhk.edu.hk. Tang: School of Business, Hangzhou City University, Hangzhou, China. Email: junjie.tang@hzcu.edu.cn. Tao: School of Entrepreneurship and Management, ShanghaiTech University, Shanghai, China. Email: taohy@shanghaitech.edu.cn. Xu: Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: jennyxu@ust.hk. We thank Ying Bai, Matilde Bombardini, David Lagakos, Yatang Lin, Qing Liu, Yifan Zhang, and seminar participants at the Chinese University of Hong Kong (Shenzhen), Shanghai University of Finance and Economics, Peking University, Tsinghua University, and Lingnan University for helpful comments and suggestions.

1 Introduction

Advancement in transportation infrastructure promotes economic development not only by lowering transportation costs efficiently but also by fostering innovation (Dong et al., 2020). Since distance hinders face-to-face communication, decreased travel time can facilitate knowledge diffusion and boost innovation. Existing literature highlights the effect of faster travel in increasing overall innovation, often measured by the total amount or quality of patents (Chen et al., 2022). However, innovations vary in their reliance on existing knowledge. Easier access to existing knowledge and old technologies may lower the cost of exploitation of known information relative to exploring unknown areas (Akcigit and Kerr, 2018).

Differentiation between the two types of innovation is essential to understanding their roles in economic development. *Incremental innovation* involves exploiting existing knowledge to make improvements within a particular "technology cluster". While such progress boosts productivity with lower risks, it eventually runs into diminishing returns (Abrams et al., 2013; Akcigit and Kerr, 2018). In contrast, *breakthrough innovation* relies on exploring unknown knowledge to create novel technology clusters, which can lead to subsequent incremental innovations. These innovations are vital to long-term economic growth, enhancing production efficiency, and inspiring future innovations (Acemoglu et al., 2022; Abebe et al., 2022).

Building upon this background, this study explores how knowledge diffusion induced by advanced transportation facilities influences firms' choice between incremental versus breakthrough innovation. To this end, we use the rapid expansion of China's high-speed rail (HSR) network over 2007-2017 as a quasi-natural experiment and examine its impact on local innovation. We construct a matched dataset composed of both information on HSR constructions and patent filings from 2001 to 2017. Our dataset covers 2.2 million Chinese patents granted in 297 Chinese prefecture-level cities, coupled with the introduction of 85 HSR lines across cities by the end of 2017.[1]

To distinguish between two types of innovations, we propose two measures to gauge the degree of breakthrough in local innovations based on the text of patent abstracts and claims, and using textual analysis techniques. First, we estimate the patent-level important index by comparing each patent's distinction from predecessors and its similarity to successors, following Kelly et al. (2021). The rationale is that breakthrough innovations tend to depart from previous studies and lead to various follow-up studies. Our novelty lies in the usage of a Large Language Model named *embedding-2*, which is developed by Zhipu AI, a leading AI company in China. LLM generates a 128-dimensional vector for the input text. We measure

[1]We stop by 2017 to evaluate the influence of each patent by comparing it with its successors in the following 3 years.

a patent's similarity to its predecessors and successors by calculating its cosine similarity to them. To make sure the comparisons are proper, we calculate the similarity for patents filed in the same technology class. Then, we sort the patents according to their importance index and calculate the share of important patents for each city and each year. Second, we focus on inter-city patent similarity, building on findings about the critical role of diversified, innovative projects in long-term innovative outcomes (Braguinsky et al., 2021).

Our analysis employs a generalized difference-in-differences (DiD) estimation strategy that exploits the quasi-random variation in HSR connections. Specifically, this DiD approach compares the changes in innovative performance in cities connected by HSR before and after the introduction of HSR, with those changes in cities without HSR connections. Under the parallel trends assumption, this approach allows us to obtain causal estimates of HSR connections on the composition of incremental versus breakthrough innovation. The parallel trends assumption requires that, in the absence of HSR connections, the composition of incremental versus breakthrough innovation would have followed similar trends for cities connected by HSR and cities lacking such connections. To verify the validity of this identifying assumption, we employ an event study specification that allows us to explicitly examine the presence of pretrends and to estimate the dynamic effects of HSR connections. Meanwhile, we recognize recent econometric concerns that staggered DiD estimates may be biased by heterogeneous treatment effects across treated cites and/or over time (Goodman-Bacon, 2021; Roth et al., 2023). We address these concerns by employing a variety of robust DiD estimators introduced in Sun and Abraham (2021), Callaway and Sant'Anna (2021), and Borusyak et al. (2024). Furthermore, to address the potential non-random assignment of HSR stations to cities, we employ the historical railway routes in 1962 as instrumental variables of HSR connections (Baum-Snow et al., 2017; Dong et al., 2020). We then adopt an instrumental variable approach, combined with long-differenced regression, to further mitigate these concerns and enhance the robustness of our estimates.

We find that better connectivity induced through HSR connections disproportionately promotes incremental innovation over breakthrough innovation, resulting in a shift of local innovation composition. Specifically, HSR connections increase the proportion of bottom 25% important patents by 3.8%, while causing a decline in the proportion of top 15% important patents by 10.9%, relative to the sample mean of these patents. Furthermore, the results of the event study specification show that innovation outcomes in HSR-connected cities and unconnected cities follow similar trends before the introduction of HSR. This finding supports the validity of the identifying assumption. We also find that the impact of HSR connectivity is initially neutral or mildly negative, yet it progressively becomes more detrimental over time. Conversely, the positive effect of HSR connectivity on incremental innovation is initially

modest but grows over time.

Our findings remain preserved when using alternative estimators and specifications. First, our estimation results are robust in the presence of heterogeneous treatment effects across treated cities and/or over time, using the estimators introduced in Sun and Abraham (2021), Callaway and Sant'Anna (2021), and Borusyak et al. (2024). Second, using the instrumental variable approach, coupled with a long-differenced regression, we demonstrate that our baseline findings are not driven by the nonrandom assignment of HSR stations. Moreover, our primary results remain statistically and qualitatively consistent under a battery of robustness checks, including controls for road and air ridership, as well as ICT development.

What explains the different effects of HSR connections on the composition of incremental versus breakthrough innovation? We show that technology spillovers induced by HSR connections shift the composition of incremental versus breakthrough innovation. Three pieces of evidence bear on this conclusion. First, we find that HSR connections enhance technology spillovers, as measured by both Jaffe's measure and Mahalanobis' measure. Second, we show that increased technology spillovers induced by HSR connections discourage breakthrough innovation, while promoting incremental innovation. Third, we show that HSR connections promote inter-city patent citations between connected city pairs, as compared to unconnected city pairs. This finding further supports the existence of knowledge spillovers across cities induced by HSR connections.

Having established that tacit knowledge spillovers are a plausible mechanism underlying the documented shift in the composition of local innovation, we further explore how technology spillovers take place across cities. Our analyses at the city-pair level show that HSR connections lead to more intercity patent transactions and licensing. We also find that the introduction of HSR fosters patent collaborations between connected cities, compared to unconnected cities, both in terms of the intensive margin and the extensive margin. These results imply that patent transactions, licensing, and collaborations are important pathways through which technology spillovers occur across cities.

Furthermore, we explore whether technology spillovers induced by HSR connections affect the direction of technological innovation, as measured by intercity patent similarity. We show that the introduction of HSR increases the patent similarity of an average city pair by 1.21 percentage points, which is equivalent to an increase of 17.90% relative to the sample mean. This finding suggests that HSR connections alter the direction of innovation between connected city pairs, leading to greater patent similarity. We further demonstrate that HSR connections enhance technology spillovers between connected city pairs relative to unconnected ones, which in turn leads to greater patent similarity between connected

cities. These results provide suggestive evidence that knowledge spillovers induced by HSR connections change the trajectory of innovation, promoting greater similarity in technological development.

Related literature. This study relates to several strands of literature. First, our study contributes to the literature on knowledge diffusion and local innovation. Innovation thrives on heterogeneous ideas, which are often generated in geographically distant places. Early work by Jaffe et al. (1993) and Keller (2002) highlight that geographic barriers impede knowledge flows, thereby constraining local innovation. Subsequent research emphasizes that improved connectivity facilitates knowledge diffusion, which increases overall innovation outputs, as measured by patent volumes or citation-based metrics (Bahar et al., 2023; Bai et al., 2024; Byun et al., 2021; Catalini et al., 2020).

A common thread in this literature is its emphasis on increased knowledge flows as a net positive for innovation quantity and quality. Our analysis extends this line of inquiry by revealing an understudied trade-off: while knowledge diffusion elevates aggregate innovation, it disproportionately incentivizes incremental innovations compared with breakthrough innovations. This reallocation of innovative effort—a dimension overlooked in prior work—suggests that the *composition* of innovation, not merely its scale, is endogenous to knowledge flow changes. By identifying this compositional shift, our study highlights the role of knowledge diffusion in shaping the direction, rather than just the intensity, of local innovations. Additionally, we explore how knowledge diffusion takes place across cities. We demonstrate that improved connectivity facilitated by HSR connections increases inter-city patent transactions, licensing, and collaborations, thereby leading to increased knowledge diffusion between connected city pairs.

The second strand of literature investigates determinants of heterogeneous innovations–specifically, factors affecting a firm's allocation of efforts between incremental and breakthrough innovations. Prior work emphasizes firm-level characteristics and incentive structures. For example, Manso (2011) and Acemoglu et al. (2022) examine how long-term compensation and job security incentivize researchers to pursue riskier, radical innovations. Central to their models is the assumption that breakthrough innovations depend less intensively on existing knowledge stocks than incremental innovations.

Our study contributes to the literature in two ways. First, we show infrastructure advancement as a novel determinant shaping the composition of two innovations. By changing the relative costs of accessing distant knowledge, improved transportation tilts effort allocation toward incremental innovations. Second, we validate the foundational assumption

in Manso (2011) and Acemoglu et al. (2022) by demonstrating incremental innovations' greater reliance on existing knowledge stock than breakthrough innovations. This evidence reinforces theoretical mechanisms that potentially drive observed compositional shifts in local innovations.

Third, our study contributes to the literature that examines the impact of transportation infrastructure on innovation. Existing studies mainly focus on the positive effects of transportation infrastructure on local innovation. For example, Agrawal et al. (2017) study the impact of roads on innovation and find that an increase in interstate highways fosters regional innovation by facilitating local within-region knowledge spillovers. Dong et al. (2020) show that the construction of HSR network increases academic publications.[2] Koh et al. (2025) find that rapid expansion of the subway in Beijing boosts patent collaborations.

While much is known about the positive impact of improved transportation infrastructure on local innovation, less is known about the influence of improved connectivity on the composition of local innovation. Our study is among the first to examine the impact of enhanced connectivity on the composition of local innovation by differentiating between incremental and breakthrough innovation using a large-scale textual analysis of Chinese patent data. Specifically, we demonstrate that HSR connections significantly alter the composition of local innovation, leading to a profound shift from breakthrough to incremental innovation.

Finally, our paper is related to the literature concerning methods for evaluating technological breakthroughs. Existing approaches fall into three categories. The first uses backward and forward citations to proxy impact (Acemoglu et al., 2022; Hall and Ziedonis, 2001), though conflating influence with knowledge diffusion limits precision. The second relies on awards or commercialization (Capponi et al., 2022), yet such metrics favor applied innovations. The third approach—textual analysis—has evolved significantly. Early work parsed patent titles and keywords (Azoulay et al., 2007), while later studies traced conceptual novelty through topic emergence (Kaplan and Vakili, 2015). Recent advances leverage computational tools: Kelly et al. (2021) quantify the extent of breakthrough for patents via term frequency–inverse document frequency (TF-IDF) algorithms that assess textual distinction from prior and subsequent work.

Our study advances this methodology by applying Large Language Models (LLMs) to patent text analysis. While Kelly et al. (2021) measure lexical divergence, we employ LLMs to

[2]Our work closely relates to Dong et al. (2020), who analyze China's high-speed rail network and its impact on academic innovation. Two distinctions emerge: First, whereas their data focus on universities, our dataset encompasses innovations across both private firms and academic institutions, capturing a more comprehensive view of local innovative activities. Second, while their findings emphasize increases in academic output quantity and quality, we demonstrate how transportation upgrades reorient innovation *composition*—shifting efforts toward incremental innovations rather than breakthrough ones.

evaluate semantic originality—capturing contextual meaning rather than keyword frequency. For example, TF-IDF may treat "dog bites human" and "human bites dog" as identical due to shared terms, whereas LLMs recognize their distinct meanings. This shift from syntactic to semantic analysis enhances precision in identifying breakthroughs.

The remainder of the paper is organized as follows. Section 2 briefly reviews the background of China's HSR construction. Section 3 details the data and key variables used in the study. Section 4 illustrates the empirical strategies. Section 5 examines the impact of HSR connections on the composition of local innovation and presents the baseline estimation results. Section 6 explores how knowledge diffusion induced by HSR connections shapes the composition of local innovation. Furthermore, Section 7 investigates if HSR connections influence the inter-city patent similarity between connected city pairs. Section 8 concludes.

2 Background of China's High-Speed Rail

China's high-speed rail (HSR) has undergone rapid expansion over 2007-2023. By the end of 2023, the total operational HSR length reached 45,000 kilometers, accounting for approximately two-thirds of the world's HSR infrastructure. This unprecedented growth provides a compelling quasi-experimental setting to study the effects of improved connectivity on knowledge diffusion and innovation. The extensive reach of the HSR system—connecting 70% of cities and facilitating mobility for over two billion residents—significantly enhances interregional knowledge spillovers by reducing travel costs and fostering face-to-face interactions among innovators.

Moreover, the construction of China's HSR network follows a centralized, national-level planning process initiated in 2004. Once approved, individual HSR projects are implemented with minimal deviations from the original blueprint. This top-down allocation mechanism ensures that the HSR introduction is largely exogenous to local economic conditions.

Rapid development. China's HSR has been developing rapidly since 2007, underpinned by strong top-down policy support and efficient implementation at both central and local levels. According to Xinhua news, the operational length of HSR tracks grew at an average annual rate of 14.4% between 2010 and 2022. During the same period, HSR ridership exhibited even more dramatic expansion, reaching 2.3 billion passengers in 2019 – representing a striking 25.8% compound annual growth rate since 2010. By 2018, HSR has become the primary mode of passenger transport in China, accounting for over 60% of China's total rail ridership.

This rapid network expansion has fundamentally transformed regional connectivity pat-

terns, particularly in eastern and central China where the HSR system has effectively bridged previous gaps in rail infrastructure (see Figure 1). These regions contain numerous large cities (with populations exceeding 500,000) situated at optimal HSR distances of 200-900 kilometers (Ollivier et al., 2014). In addition to the national HSR grid, high-speed intercity rails enhance regional connectivity within provincial boundaries through additional intermediate stops.

The unprecedented scale and speed of China's HSR development, coupled with its profound integration effects among cities, creates an ideal natural experiment for investigating the relations between accelerated knowledge diffusion and local innovation.

Top-down planning. The rapid expansion of China's HSR network has been fundamentally enabled by the country's centralized planning process. The development follows a strict national blueprint – the Medium- and Long-Term Railway Plan (MLTRP) formulated in 2004 through inter-ministerial coordination among three key central government agencies. This master plan established a strategic roadmap for HSR development towards 2020, with its centerpiece being the construction of a national " four vertical and four horizontal" HSR grid spanning 12,000 kilometers. The government significantly expanded these ambitions in 2008, raising the target for passenger-dedicated lines (PDLs) to 16,000 kilometers by 2020.[3]

The MLTRP's status as a binding national policy directive ensures remarkable implementation consistency across regions. Local deviations from the approved project plans have been exceptionally rare(Lawrence et al., 2019). This centralized decision-making process creates a quasi-experiment, where HSR network expansion can be treated as largely exogenous to local economic conditions including the innovative activities of firms and universities. This institutional setting provides a unique opportunity to examine how transport infrastructure shocks affect local innovative choices.

Innovation and HSR. China's HSR expansion has coincided with a remarkable acceleration in innovative activities nationwide. According to our patent dataset, the average city-level patent filings increased from 380 in 2008 to 1,150 in 2016, representing a compound annual growth rate of 15%. Figure 2 illustrates this transformation by comparing the geographic distribution of cities with above-average patent fillings (exceeding 550, our sample mean) before and after HSR introduction. Notably, the newly increased 20 cities between 2008 and 2016 (shown as blue dots in Panel b) exhibit greater geographic dispersion while demonstrating strong spatial correlation with newly constructed HSR routes (represented by red lines). This

[3]As the original 2004 targets were achieved ahead of schedule by 2015, the MLTRP was revised in 2016 to establish an even more ambitious "eight vertical and eight horizontal" network spanning 38,000 kilometers by 2025.

pattern suggests a potentially significant relationship between HSR connectivity and local innovations.

A deep examination of patent data implies that the effects of HSR are potentially associated with facilitated inter-city knowledge flows. Figure 3 reveals a striking overlap between the HSR network (red lines) and inter-city patent collaboration linkages (grey lines), with collaboration density visibly higher among HSR-connected city pairs. These observed correlations motivate our investigation into the role of knowledge diffusion in China's recent innovation boom and the shift of innovation composition.

3 Data and Measurements

3.1 Datasets

We begin our analysis by constructing a matched dataset with information on high-speed railway construction and patenting across cities in China from 2001 to 2017. This dataset is compiled from four different sources, enabling us to conduct a comprehensive study on the effect of the construction of HSR on innovation in China.

The first dataset consists of Chinese patent data from the China National Intellectual Property Administration (CNIPA). The Chinese patent data covers all the invention patent filings and grants between 1985 and 2023, detailing an applicant's bibliography information, filing date, grant date, abstract, claims, and references (patents) cited. Based on the bibliographic information of applicants, we construct measures of innovation activities for each prefecture-level city in China. Moreover, by using patent abstracts and claims, we construct measures of patent quality and knowledge spillover, which are discussed in detail in the following section.

The second dataset consists of the construction and operation of HSR in China from 2008 to 2017 provided by the Ministry of Railway (MOR). HSR is defined as railways that operate at 250 - 350 km/h. The MOR announced its HSR plan in November 2006, and 1121 HSR stations and 135 HSR lines were in operation till 2020.

The third dataset consists of the city-level statistics of prefecture-level cities in China. We collect the statistics, such as GDP, employment, fiscal expenditure, and investment, from the *China City Statistical Yearbook* from 2001 to 2017. Please note that we only study the innovative activities of the prefecture-level cities and the four municipalities. Although the prefectures, leagues, and autonomous prefectures and counties are equivalent to the prefecture-level cities in administration, they are not included in our analysis.

We merge the three batches of data based on city names. This process results in samples at the city-year and city-pair-year levels. These two samples cover 2.2 million granted Chinese patents across 297 Chinese prefecture-level cities from 2001 to 2017. These cities are connected by 95 HSR lines by the end of 2017. We choose 2017 as the cutoff year to allow evaluation of each patent's influence by comparing it with its successors over the following 3 to 5 years.

3.2 Key variables

HSR Connection. By using the time of operation of HSR stations and HSR lines, we construct two types of HSR connection indicators for prefecture-level cities. First, based on the time of operation of each HSR line, we identify the earliest year that a city i is connected by HSR. Second, based on the evolution of the HSR network, we identify the earliest year that two cities, city i and city j, are connected by the HSR network, both directly and indirectly. For instance, they are connected if city i and city j are located in the same HSR line. Meanwhile, even if city i and city j are located in two different HSR lines, they are also considered connected by HSR if the two HSR lines intersect in city k.

Patent Importance. In the baseline analysis, we construct an importance index for each patent following the approach proposed by Kelly et al. (2021). However, unlike Kelly et al. (2021), we enhance this method by incorporating semantic information through the use of a Large Language Model (LLM) during the text embedding process, rather than relying on the traditional TF-IDF method. Notably, the LLM offers several advantages over TF-IDF in the context of text embedding. First, the LLM captures semantic meanings within the text, whereas TF-IDF is purely based on term frequency and does not account for contextual meaning. Second, TF-IDF generates high-dimensional, sparse vectors, which can introduce noise and increase the computational burden for downstream tasks. Third, TF-IDF embeddings are highly dependent on the specific wording of the input text, which may lead to biased results. In contrast, LLM-based embeddings are derived from fixed model parameters, thereby reducing sensitivity to specific phrasing and mitigating potential bias.

To begin with, we obtain the embedding of each patent abstract using the Large Language Model *embedding-2* developed by Zhipu AI, which generates a 128-dimensional vector to represent the input text in the vector space.[4]

Next, we calculate the indicator of patent importance for each patent by comparing its semantic similarity to both its predecessors and its successors, as defined in Equation

[4]We also construct the text embedding with E-5, an open-source Large Language Model developed by Microsoft and our results are robust to alternative embedding models.

(1) (Kelly et al., 2021). The patent importance indicator identifies patents that exhibit substantial novelty relative to their predecessors and exert significant influence on subsequent innovations.

$$\text{Patent Importance Index}_{p,t} = \frac{CosSim_{p,t,+3}}{CosSim_{p,t,-3}} = \frac{\underset{p' \in x, s \in [t+1,t+3]}{avg} CosSim(\vec{v}_{p,t}, \vec{v}_{p',s})}{\underset{p' \in x, s \in [t-3,t-1]}{avg} CosSim(\vec{v}_{p,t}, \vec{v}_{p',s})} \tag{1}$$

Specifically, for patent p filed in year t, we calculate its average similarity between the patents that share the same technology class and are filed between year $t-3$ and $t-1$. Similarly, we also calculate the average similarity between patents that share the same technology class and are filed between years $t+1$ and $t+3$. A higher value of the patent importance index signifies a patent that is both novel and influential. Intuitively, a patent that exhibits high similarity to its successors but low similarity to its predecessors can be considered both novel and influential, and is thus regarded as a breakthrough innovation. Thus, the patent importance index allows for the identification of breakthrough patents, which are characterized by their substantial novelty and significant influence on subsequent technological developments.

Then, we calculate the share of important patents for each city (Acemoglu et al., 2017). Specifically, We first identify patents that are above the p^{th} percentile of the distribution based on *patent important index*. Second, we calculate city i's share of important innovations for each year as defined in Equation (2).

$$\text{Top Important Share}_{i,t} = \frac{z_{it}(p)}{z_{it}(0)} \tag{2}$$

Intuitively, a higher *Top Important Share* indicates that city i undertakes more *important patents*, reflecting the composition of local innovation. In the baseline analysis, we classify the top 15 percent, 20 percent, and 25 percent based on their patent importance index as breakthrough innovation. Conversely, patents in the bottom 25 percent, 30 percent, and 70 percent are categorized as incremental innovation.

Technology Spillover. We measure the pool of technology spillovers based on the methodology proposed by Jaffe (1986) and Bloom et al. (2013). The strength of technology spillover from city j to city i depends on two aspects. On the one hand, the level of knowledge stock in city j affects the strength of technology spillover. On the other hand, the proximity of technology, i.e., the similarity of structures of existing technologies between two cities, also

affects the strength of technology spillover (Bloom et al., 2013).

We first construct the measures of technological proximity following Jaffe (1986). Jaffe's measure technological proximity assumes that spillovers only occur within the same technology class, which rules out spillovers between different classes. Specifically, Jaffe's measure of technological proximity between city i and city j is specified as follows:

$$tech_{ijt}^{Jaffe} = \frac{U_{it}U'_{jt}}{||U_{it}|| \cdot ||U_{ij}||} = \frac{\sum_{x=1}^{N} U_{i,t}U_{j,t}}{\sqrt{\sum_{x=1}^{N} U_{i,t}^2} \cdot \sqrt{\sum_{x=1}^{N} U_{j,t}^2}} \tag{3}$$

where $U_{i,t} = (u_{i1,t}, u_{i2,t}, ..., u_{i\tau,t}, ..., u_{iN,t})$ is a vector that denotes city i's proportion of patents in each technology class τ at 3-digit IPC level.

Third, we construct the pool of technology spillovers for city i in year t, using Equation (4). Intuitively, the summation of technology spillovers from all other cities to city i captures the strength of technology spillover faced by city i in year t.

$$\text{SpillTech}_{i,t}^{Jaffe} = \sum_{i \neq j} tech_{ij,t}^{Jaffe} Stock_{j,t} \tag{4}$$

where $Stock_{j,t}$ is the stock of knowledge in city j, which is calculated with a perpetual inventory method with a $\delta = 15\%$ depreciation rate. [5] Here, we use the number of patents to proxy for the level of knowledge stocck. A higher value of $SpillTech_{i,t}^{Jaffe}$ indicates a greater amount of technology spillovers.

Third, we construct a modified measure of technology spillovers, denoted as $SpillTech_{i,t}^{Maha}$, based on the Mahalanobis distance of technological proximity, as proposed by Bloom et al. (2013).[6] This Mahalanobis measure captures technology spillovers both within and across technology classes. The calculation of this Mahalanobis measure of technology spillovers, $SpillTech_{i,t}^{Maha}$, is presented in Appendix A.

City-Pair Level Statistics. We further construct four city-pair-level statistics on innovation activities to support our analysis.

First, we construct patent citations from city j to city j in year t, which measure the flows of knowledge from city j to city i.

[5] Specifically, knowledge stock is computed using the equation: $Stock_{j,t} = (1-\delta)Stock_{j,t-1} + Patent_{j,t}$. Intuitively, the stock of knowledge of city j in year t is the number of newly filed patents plus the depreciated previous patent stock. The patent system in China began in 1985, and we can precisely estimate the capital stock for all cities from 1985.

[6] Relative to Jaffe distance of technological proximity, Mahalanobis distance of technological proximity provides a more comprehensive assessment of technological relatedness

Second, we calculate the number of collaborative patents filed by applicants from city i and city j in year t. The patent data from CNIPA provides the names of all applicants in the patent filing and the address of the first applicant. However, the patent data from CINPA fails to provide the address of the second applicant. To solve this problem, we identify the address of the second applicant from the administrative data of firm registration. Therefore, we can identify all the collaborative patents between city i and city j. Additionally, for each city pair in year t, we identify whether the research teams involved in collaborative patents are newly formed or pre-existing. Based on this, we compute both the total number of research teams engaged in collaborative patents and the number of newly formed teams.

Third, we calculate the number of patent transactions sold from city j to city i. As the authority requires, each patent transaction must be registered in the CNIPA. Thus, we identify the buyer and seller for each patent transaction, as well as their locations, and construct the number of patents sold from city j to city i each year. Similarly, for each patent licensing, we identify the patent licensor and licensee, along with their respective locations. We then compute the number of patents licensed from city j to city i.

Finally, we construct a measure of technology similarity between patents in city i and city j. We calculate the pairwise similarity between patent vectors of city i and city j in each year and for each technology class. The average similarity between city i and city j in year t is defined in Equation (5).

$$\text{InnovSim}_{ij,t} = \underset{p \in i,\ p' \in j}{Avg\ Pairwise\ CosSim}(\vec{v}_{p,i,t}, \vec{v}_{p',j,t}) \tag{5}$$

4 Empirical Framework

4.1 Baseline model specification

The main goal of this paper is to identify the impact of HSR connectivity on the composition of incremental versus breakthrough innovation. However, several endogeneity concerns may plausibly bias the estimation of the causal impact of HSR connections on the composition of local innovation. For example, reverse causality may arise because highly innovative cities are more likely to receive more HSR investment. Additionally, omitted variable bias may be another concern as more developed cities may not only attract HSR projects but also benefit from greater innovation investment due to their better access to infrastructure.

To obtain credible estimates, leveraging the staggered rollout of HSR connections across

cities, we use a generalized difference-in-differences (DiD) strategy to examine the effects of HSR connections on the composition of incremental versus breakthrough innovation at the city level. This DiD strategy compares the changes in the composition of local innovation in connected cities (i.e., treatment group) before and after HSR opening with those in cities that do not receive HSR connections (i.e., control group).

As a baseline specification, we use the following two-way fixed effect (TWFE) specification at the city level:

$$y_{it} = \alpha + \beta HSR_{i,t-1} + \lambda' X_{it} + \tau_i + \delta_t + \epsilon_{ijt} \tag{6}$$

where y_{it} measures the innovative activities, including measures that capture the composition of incremental versus breakthrough innovation, in city i in year t. $HSR_{i,t-1}$ takes a value of 1 if city i is connected to any HSR line in year $t-1$ and 0 otherwise. Here, we use the lagged HSR connections because innovative activities are unlikely to adjust instantaneously. X_{it} are a vector of time-varying characteristics at the city level, including GDP per capita, average wage, percentage of GDP from the secondary industry, and the percentage of GDP from service sector. τ_i is the city fixed effect; δ_t is the year fixed effect. We cluster standard errors at the city level. The coefficient of interest is β, which captures the effects of HSR connections on the composition of incremental versus breakthrough innovation.

4.2 Event Study Specification

When estimating the effects of HSR connections on the composition of local innovation, the validity of the baseline DID model relies on the identifying assumption that, in the absence of HSR connections, the composition of local innovation of HSR-connected cities and that of unconnected cities would have similar trends.

To verify the validity of the parallel trends assumption, we employ an event-study version of the TWFE model. Specifically, we use the following specification:

$$y_{it} = \alpha + \sum_{k=-7,\ k\neq -1}^{7} \beta_k D_{i,t-k} + \lambda' X_{it} + \tau_i + \delta_t + \epsilon_{ijt} \tag{7}$$

where y_{it} measures the composition of local innovation in city i in year t. $D_{i,t-k}$ is a set of indicator variables that take the value of one if the HSR is first introduced in city i in year $t-k$. Here, k captures the time period relative to the initial year of HSR connections in city i. The reference category is one year before the introduction of HSR. The coefficient β_k, for $k<0$, captures the pre-trends in the composition of local innovation performance prior to

the introduction of HSR. The coefficient β_k, for $k \geq 0$, captures the dynamic effects of HSR connections on the composition of local innovation following its introduction. If the parallel trends assumption holds, we expect that the estimated coefficients prior to the introduction of HSR (i.e., β_k for $k < 0$) are all statistically close to zero.

One may be concerned that the TWFE model estimated using OLS may be biased in the presence of heterogeneous treatment effects across treated cities and/or over time (Goodman-Bacon, 2021; Roth et al., 2023). To address this concern, we explore the existence of pretrends by estimating the event-study version of the TWFE model using a set of robust estimators introduced in Sun and Abraham (2021), Callaway and Sant'Anna (2021), and Borusyak et al. (2024). By eliminating "bad" comparisons between newly treated and already treated cities, these estimators provide consistent estimates of the effects of HSR connections on the composition of local innovation, even when treatment effects vary across treated cities and/or over time.

4.3 Long-differenced Regression: Instrument Variable Approach

There is a concern that central governments may place HSR stations in booming cities where those cities have more investment in local innovation. Thus, innovative cities are more likely to be connected to HSR networks. To address this endogeneity issue, we employ an instrumental variable (IV) approach. Following Baum-Snow et al. (2017) and Dong et al. (2020), the city-level instrument we construct is a dummy variable indicating whether city i was connected by historical railway networks in 1962. Similarly, we construct the corresponding city-pair level instrument, which is a dummy variable of whether city i and j are connected by the historical railway networks.

Our instruments are valid under the following two conditions: (i) the 1962 historical railway networks positively predict the modern HSR networks (or the relevance condition). (ii) the historical railway networks do not affect local innovation through channels other than HSR connections (or the exclusion restriction condition). The relevance condition can be confirmed by the first-stage F-statistics. As the locations of 1962 historical railway routes are unlikely to be correlated with the local investment in innovation, our instruments may satisfy the exclusion restriction condition. As our instruments are time-invariant, we can only employ long-difference regressions when conducting the IV approach. The long-difference regressions give relatively more weight to persistent changes in local innovation while they reduce the size of the sample.

5 The Impact of HSR Connections

5.1 Baseline Results

Table 1 reports the estimated effects of HSR connections on the composition of local innovation. We differentiate innovations based on their nature of exploitation versus exploration and study the effects of HSR connections on the proportion of incremental and breakthrough innovation, respectively. Columns (1)-(3) present the estimation results for breakthrough innovation, in which the dependent variables are the proportion of top important patents located in the top 15th, 20th, and 25th percentiles, respectively. The coefficients of $\text{HSR}_{i,t-1}$ are negative and statistically significant, showing that the HSR connections have sizable negative impacts on the share of the most important patents. These results suggest that HSR connections discourage breakthrough innovation, which relies on local innovators exploring uncharted areas.

Columns (4)-(6) provide the estimation results for incremental innovation, whereby the dependent variables correspond to the proportion of bottom important patents falling within the bottom 25th, 30th, and 70th percentiles, respectively. The estimated coefficients of $\text{HSR}_{i,t-1}$ are positive and statistically significant. The positive coefficients of $\text{HSR}_{i,t-1}$ suggest that HSR connections foster incremental innovation, which relies on the exploitation of existing knowledge. Taken together, these results imply that HSR connections promote incremental innovation but discourage breakthrough innovation, leading to a compositional shift in local innovation from breakthrough innovation to incremental innovation.

The effects of HSR connections on the composition of incremental versus breakthrough innovation are economically significant. For example, according to column (1) of Table 1, HSR connections are associated with a 10.9% decrease in the proportion of breakthrough patents (i.e., the top 15% important patents), relative to the sample mean proportion of such patents. On the other hand, based on column (4), HSR connection increases the proportion of incremental patents (i.e., the bottom 25% important patents) by 3.8% (relative to the sample mean). These findings consistently suggest that HSR connections induce local innovators to focus on exploiting existing knowledge rather than exploring unknown areas, thereby reshaping the composition of local innovation.

A key assumption for the generalized DID design to identify the causal effects of HSR connections on the composition of local innovation is that innovation outcomes in the treatment and control cities would have followed parallel trends in the absence of HSR connections. Although this identifying assumption is not directly testable, we can provide supporting evidence to its validity by testing the presence of pre-trends before the introduction of HSR

connections. Figure 4 provides the event study graph for the effects of HSR connections. As shown in Figures 4(a) – 4(c), we find no significant differences in the top important patents between connected and unconnected cities in the years leading up to the HSR connections. Moreover, we observe negative and significant effects on the top important patents four years following the introduction of HSR connections. On the other hand, as Figures 4(d)–4(f) show, we observe little difference in the bottom important patents between treated and control cities prior to the introduction of HSR connections. After the introduction of HSR connections, we also document clear increases in the bottom important patents in connected cities relative to the unconnected cities.

Overall, these findings support the validity of our identifying assumption that the proportions of incremental and breakthrough innovation in connected and unconnected cities follow common trends before the introduction of HSR connections. Additionally, these graphs also shed light on the dynamic effects of HSR connectivity. Specifically, HSR connectivity has a neutral or mildly negative effect on breakthrough innovation in the initial years, which gradually evolves into a more pronounced detrimental impact. In contrast, the positive impact on incremental innovation, though modest at first, accumulates and becomes significantly stronger over time.

Figure 5 presents the graphs from the event study using a set of alternative DID estimators that are robust to heterogeneous treatment effects (Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; Borusyak et al., 2024). These graphs uniformly show that the estimated coefficients before the introduction of HSR are statistically close to zero, indicating that there are no discernible pretrends. These graphs further confirm the dynamic effects of HSR connections: the negative effects of HSR connections on breakthrough innovation intensify over time, whereas the positive effects on incremental innovation become increasingly pronounced.

5.2 Robustness checks

IV Approach in Long-differenced Regressions. Table 2 presents the long-difference regression results at the city level, where the instrumental variable of HSR connections is the historical railroad in 1962. As shown in columns (1)–(3) of Table 2, the estimated coefficients of the instrumented HSR connections variable are consistently negative and statistically significant in column (3), albeit imprecisely estimated in columns (1) and (2). These results further confirm our main findings that HSR connections discourage breakthrough innovation among connected cities, relative to unconnected cities. Moreover, columns (4)–(6) consistently demonstrate that HSR connections have positive effects on the proportion of bottom important patents. While the estimated coefficients for HSR connections in columns

(4)–(5) are imprecisely estimated, the overall pattern remains consistent. Nevertheless, these results should be interpreted with caution, as they are based on a small number of observations."

Placebo Tests We conduct placebo tests to examine whether our baseline findings have been driven by chance. Specifically, we randomly assign cities launching HSR, set a pseudo year for HSR opening, and then replicate the baseline regressions. Figure 6 reports the corresponding results for the proportion of top and bottom important patents, respectively. Since the estimated coefficients of pseudo-HSR connections are centered around zero, these results provide little evidence that unobserved inherent differences between treatment and control groups trigger our baseline results.

Road and Air Ridership As the expansion of road infrastructure and airports can reduce inter-city transportation costs and increase communication among inventors, it is possible that the observed changes in the composition of local innovation may be driven by the expansion of roads and airports rather than the HSR connections(Lin, 2017; Lin et al., 2021; Dong et al., 2020). To alleviate this concern, we add road and air ridership as additional controls. As shown in Table 3, our main results remain statistically significant after controlling for both road and air ridership. These findings further reinforce the conclusion that HSR connections promote incremental innovation while impeding breakthrough innovation, thereby resulting in a structural change in local innovation.

ICT Development The rapid development of the Internet and other Information and Communication Technologies (ICT) can enhance communication among innovators across cities, potentially diminishing the impact of HSR connections on local innovation (Agrawal et al., 2017). To explore this issue, we collect city-level data on the adoption of ICT and use the number of broadband subscribers as a proxy for ICT development. We present the regressions that include this ICT measure in Table 4. Our main findings remain robust. The estimated coefficients of HSR connections are statistically and quantitatively similar to those obtained in the baseline model. However, ICT development has minimal and statistically insignificant effects on the composition of incremental versus breakthrough innovation. These findings indicate that face-to-face interactions, enabled by HSR connections, play a more significant role in shaping the composition of local innovation than online communication facilitated by the advancement of ICT.

6 Mechanism: The Role of Technology Spillovers

To explore the mechanisms behind HSR's differential effects on incremental and breakthrough innovation, we first propose a conceptual framework to analyze the changes in the relative costs of pursuing each type of innovations after HSR introduction. Guided by the conceptual model, we empirically examine the role of knowledge spillovers in shifting the composition of local innovation. We also document the existence of knowledge spillovers across cities, as measured by inter-city patent citations. Moreover, we examine how knowledge spillovers take place across cities, utilizing information on inter-city patent transfers, patent licensing, and patent collaborations.

6.1 A Conceptual Model of Innovation

We outline the conceptual model and core mechanisms verbally here, with formal modeling details provided in Appendix B. Consider two geographically distant cities, each with a representative firm allocating resources between incremental and breakthrough innovation. Innovators in each city produce using two inputs: (i) knowledge (both local and distant) and (ii) local capital (such as scientists, laboratories), combined via Cobb-Douglas aggregation.

The innovation types differ fundamentally in their input composition. Incremental innovation is knowledge-intensive, relying heavily on available knowledge stock. On the contrary, breakthrough innovation is less knowledge-intensive but yields higher returns than incremental innovation.

As the input, knowledge is a homogeneous public good, which is free and non-excludable. However, its accessibility decays with distance. Innovators can fully use local knowledge but only access a fraction of distant knowledge due to geographic barriers. Capital inputs are costly to innovators, as they have to pay wages to scientists and establish laboratories. For simplicity, we assume a fixed capital endowment for each city.

Given the characterization of the model, the advancement in transportation infrastructure (such as HSR connections) generates two effects. By reducing the geographic decay of knowledge, it enables innovators to access a larger portion of distant knowledge. This expanded knowledge access, combined with the knowledge-intensive nature of incremental innovation, leads firms to shift their efforts toward incremental projects (as formalized in Proposition 1 of Appendix B). An increased knowledge abundance lowers the relative cost of incremental innovation.

Moreover, the facilitated knowledge flows produce a convergence effect. As transportation improves, the mix of innovative inputs (both knowledge and capital) becomes more similar

across cities. Innovators in different places are likely to undertake more similar types of innovation (Proposition 2, Appendix B).

In the following part of this section, we focus on knowledge spillovers, which is directly measured by inter-city patent citation, and empirically examine their roles in shifting local innovation composition. In Section 7, we discuss the direction of innovation for different cities, corresponding to the convergence effect.

6.2 The Role of Technology Spillovers in Shaping Innovation

To explore the role of technology spillovers in shaping local innovation, we begin our analysis by investigating if HSR connections induce cities's exposure to technology spillovers, as measured by Jaffe's distance and Mahalanobis' distance. Specifically, we separately regress Jaffe's and Mahalanobis' measures of technology spillovers on the dummy variable of HSR connections. Table 5 reports the regression results, in which the measures of technology spillovers are discounted at the 5% depreciation rate in columns (1)–(2) and at the 15% depreciation rate in columns (3)–(4), respectively. The coefficients of HSR connections are positive and significant across all columns, showing that, compared to unconnected cities, HSR connections enhance technology spillovers received by HSR-connected cities. These findings further suggest that HSR connections reshape the composition of local innovation by facilitating increased exposure to external knowledge flows.

Having established that HSR connections promote technology spillovers, we then study if increased knowledge spillovers induced by HSR connections influence the composition of local innovation. In Panel A of Table 6, we separately regress the proportions of top and bottom important patents on Jaffe's measure of technology spillovers. The estimated coefficients of the HSR connections in columns (1)–(3) are negative and significant, showing that technology spillovers discourage breakthrough innovation that explores unknown areas. In particular, a 10% increase in Jaffe's measure of technology spillovers is associated with a 1.4% decrease in the proportion of the top 15% important patents (relative to sample mean).[7] On the other hand, the coefficients of the HSR connections are positive in columns (4)–(6) and statistically significant in column (3), although they are imprecisely estimated in columns (5)-(6). These results suggest that knowledge spillovers are more likely to stimulate incremental innovation, which builds on existing knowledge. In sum, we find that knowledge spillovers facilitated by HSR connections promote incremental innovation, while inhibiting breakthrough innovation.

In addition, in Panel B of Table 6, we separately regress the proportions of top and

[7]The sample mean of the proportion of the top 15% important patents is 0.115.

bottom important patents on Mahalanobis' measure of technology spillovers. The estimated results reveal a consistent pattern: knowledge spillovers induced by HSR connections reduce the proportion of breakthrough innovation, while these knowledge spillovers increase the proportion of incremental innovation.

Taken together, these results suggest that knowledge spillovers induced by HSR connections facilitate the exploitation of existing knowledge, which is beneficial to incremental innovation. However, these knowledge spillovers have detrimental effects on the exploration of unknown areas, thereby constraining breakthrough innovation. As a result, HSR connections shape the decomposition of local innovation by promoting external knowledge spillovers that favor incremental over breakthrough innovation.

As suggested by the theoretical framework of Manso (2011), individuals and organizations face trade-offs between incremental and breakthrough innovation when making innovation decisions. In the presence of technology spillovers, individuals and organizations are more exposed to existing knowledge, which reduces the cost of undertaking incremental innovation compared to that of conducting breakthrough innovation (Byun et al., 2021). Therefore, our findings align with the argument in the literature that technology spillovers promote innovation that relies on the exploitation of existing knowledge, while discouraging innovation that involves the exploration of unknown areas (Manso, 2011; Byun et al., 2021).

6.3 Existence of Knowledge Spillovers: Evidenced by Patent Citations

Since a patent represents a piece of previously existing knowledge, patent citations are usually considered as a good way to measure knowledge flows, especially technological knowledge flows. In addition, knowledge flows across cities take place when one city cites another city's patents (Jaffe et al., 1993). Therefore, to further explore the existence of knowledge spillovers, we examine the impact of HSR connections on inter-city patent citations.

To study the effects of HSR connection on technology spillovers across cities, we conduct our analysis at the city-pair level. The econometric specification at the city-pair level is as follows:

$$y_{ijt} = \alpha + \gamma HSR_{ij,t-1} + \lambda' X_{ijt} + \tau_{ij} + \delta_{it} + \delta_{jt} + \epsilon_{ijt} \tag{8}$$

where y_{ijt} denotes the technology spillovers from city j to city i in year t, as measured by inter-city patent citations, patent transfers, patent licensing, and patent collaborations respectively. Specifically, we use patent citations, patent transfers, patent licensing, and

patent collaborations between city i and city j as the dependent variables, respectively. $HSR_{ij,t-1}$ takes a value of 1 if city i is connected to city j by any HSR line in year $t-1$ and 0 otherwise. X_{ijt} are a vector of time-varying characteristics at the city-pair level. τ_{ij} is the city-pair fixed effects. δ_{it} is citing-city by year fixed effects, and δ_{jt} is cited-city by year fixed effects. Standard errors are clustered at the city-pair level. The coefficient of interest, γ, estimates the effects of the HSR connections on inter-city knowledge diffusion.

In columns (1)–(2) of Table 7, we regress inter-city patent citations on the dummy variable of HSR connections, in which the dependent variables are the number of patent citations and the number of university patent citations between city pairs, respectively. As shown in column (1), the estimated coefficient of HSR connections is positive and significant, showing that HSR connections increase the number of patent citations between connected city pairs by 19.8%, compared to unconnected city pairs. We also find that, after HSR connections, the number of university patent citations between HSR-connected city pairs increases by 14.4%, as shown in column (2). Taken together, these results demonstrate that HSR connections promote patent citations between connected city pairs, compared to unconnected city pairs. This finding provides compelling evidence for the existence of technology spillovers between cities induced by HSR connections.

6.4 How Technology Spillovers across Cities Take Place?

Having established that technology spillovers are a plausible mechanism driving the documented shift in the composition of local innovation, we further explore how technology spillovers take place across cities. It is well established that HSR connections substantially reduce travel time and costs, thereby facilitating the interactions and communications among individuals and organizations in different cities (Lin, 2017; Hanley et al., 2022). These enhanced interactions foster increased patent collaborations, transactions, and licensing—key channels through which technological spillovers materialize. Accordingly, we examine three specific dimensions of technology spillovers. First, we assess if HSR connections influence inter-city patent transactions. Second, we analyze if HSR connections facilitate patent licensing across cities. Third, we explore if HSR connections promote greater inter-city patent collaborations.

Inter-city Patent Transactions Patent transactions, which involve contractual agreements enabling buyers in one city to acquire patent-protected knowledge from assignors in another, serve as a critical channel for knowledge spillovers(Liu et al., 2022). In particular, following a patent transfer, close interpersonal contact and intensive communication between

the inventor and the assignee help to reduce information asymmetries, thereby facilitating the exchange of both explicit and tacit knowledge(Maurseth and Svensson, 2020). To explore this issue, we examine whether HSR connections promote inter-city patent transactions as a pathway for knowledge spillovers. The regression results are presented in column (3) of Table 7. The estimated coefficient of the HSR connection variable is positive and significant, showing that HSR connections between a given city pair increase patent transactions between these cities by 6.36%, compared to unconnected city pairs. These results suggest that HSR connections enhance knowledge spillovers through increased patent transactions across HSR-connected city pairs.

Inter-city Patent Licensing The licensing of patents between inventors in one city and users in another city also serves as a complementary means of technology spillovers(Nelson, 2009). Specifically, by purchasing the right to use a patent, licensees can require patent holders to provide additional technical materials and in-depth technical assistance, thereby facilitating a smooth technology transfer.[8] Thus, we further examine the ways through which knowledge spillovers occur by testing if HSR connections promote inter-city patent licensing. Column (4) of Table 7 presents the regression results. The estimated coefficient of the HSR connections is positive and significant, showing that HSR connections increase inter-city patent licensing by 1.18%. This finding implies that patent licensing is an important channel through which inter-city technology spillovers, induced by HSR connections, take place.

Inter-city Patent Collaborations As HSR connections can reduce the costs of interacting with distant innovators, they may encourage innovators to both establish new and strengthen existing inter-city collaborations. Moreover, since such collaborations involve direct interaction as well as complex, nuanced communication and coordination, they can, in turn, facilitate technology spillovers between connected cities(Boudreau et al., 2017). Therefore, to examine whether HSR connections foster knowledge spillovers through this collaborative channel, we analyze their impact on both the intensive and extensive margin of inter-city patent collaborations.

We begin our analysis by testing whether HSR connections affect inter-city collaboration at the intensive margin, in which the dependent variables are the number of collaborative patents and the number of collaborative research teams at the city-pair level, respectively. Columns (1) and (2) of Table 8 present the estimation results. We find that HSR connections increase

[8]In addition to patent documents, patent licensing agreements typically include clauses specifying the supplementary technical materials and assistance that the patent holder must provide. These terms may include detailed laboratory reports, user manuals, on-site guidance, and technical training, which ensures that the licensee has easier access to explicit knowledge, as well as tacit knowledge.

both the number of collaborative patents and the number of collaborative research teams between connected city pairs. Specifically, following the introduction of HSR connections, the number of collaborative patents between connected city pairs increases by 7.19%, compared to unconnected city pairs. Concurrently, the number of collaborative research teams between these city pairs rises by 5.27%. These results show that HSR connections enhance patent collaborations at the intensive margin between pairs of cities connected with HSRs.

We then examine the effects of HSR connections on the extensive margin of inter-city patent collaborations. Columns (3) and (4) of Table 8 present the estimation results, where the dependent variables are the number of new collaborative research teams and the number of patents filed by these newly formed teams, respectively. Specifically, the result in columns (3) shows that HSR connections lead to a 4.46% increase in the number of new collaborative research teams. As shown in column (4), we also find that HSR connections increase the number of collaborative patents filed by these new teams by 6.01%. These results indicate that HSR connections also enhance inter-city patent collaborations at the extensive margin.

Overall, the above results show that HSR connections increase patent collaborations between HSR-connected cities at both the intensive and extensive margins. These findings also demonstrate that, in the presence of HSR connections, local innovators are more likely to establish new and more collaborations with distant innovators. Given that patent collaborations facilitate interactions and communications among individuals and organizations, they also provide easier access to knowledge, particularly tacit knowledge, in collaborative cities. Thus, these results suggest that patent collaborations serve as a crucial conduit for knowledge spillovers across geographic boundaries.

7 Extension: Direction of Innovation across Cities

So far, we have documented that HSR connections lead to technology spillovers, which in turn discourage breakthrough innovation while encouraging incremental innovation. However, much less is known about whether technology spillovers induced by HSR connections affect the direction of innovation across cities, as measured by inter-city patent similarity.

First, we study if HSR connections alter the trajectory of innovation between connected city pairs. The pairwise structure of inter-city patent similarity also allows us to conduct a city-pair level analysis to study the impact of HSR connections on the technological direction of innovation between city pairs. Specifically, we use the patent similarity between city i and city j as the independent variable of interest and perform our analysis at the city-pair level using the specification detailed in Equation (8). In this case, the coefficient of interest, γ,

measures the impact of HSR connections on the direction of technological innovation between city pairs.

Columns (1)-(2) of Table 9 present the estimation results, showing that the estimated coefficients of $HSR_{ij,t-1}$ are positive and statistically significant. Specifically, column (2) shows that the HSR connections increase an average city pair's patent similarity by 1.21 percentage points, which is equivalent to a 17.90% increase in inter-city patent similarity (relative to the sample mean).[9] These results demonstrate that the introduction of HSR connections shifts the trajectory of innovation between connected city pairs toward a more similar direction.

Furthermore, using the historical railroad in 1962 as an instrument of HSR connections, we perform long-differenced regressions and report these regression results in Table 10. Specifically, column (1) of Table 10 reports the 2SLS regression results, showing that HSR connections have a positive and significant impact on inter-city patent similarity. Column (2) shows the first-stage regression results, suggesting that the instrument satisfies the relevance condition and providing suggestive evidence of the validity of using the historical railroad in 1962 as an instrument. These long-differenced regression results further confirm that HSR connections lead to greater similarity between connected city pairs, thereby shaping the direction of innovation between HSR-connected cities.

Second, we explore how HSR connections affect the direction of innovation across cities. In particular, we begin our analysis by investigating if HSR connections lead to cross-city technology spillovers between connected city pairs. To explore these issues, we regress the inter-city technology spillovers as measured by Jaffe's and Mahalanobis' measures on the dummy variable of $HSR_{ij,t-1}$. As shown in columns (1) –(2) of Table 11, the estimated coefficients of $HSR_{ij,t-1}$ are positive and significant, showing that HSR connections enhance technology spillovers between connected city pairs, compared to unconnected city pairs.

We further explore if inter-city technology spillovers influence the direction of innovation across cities. Specifically, we regress inter-city patent similarity on inter-city technology spillovers as measured by Jaffe's and Mahalanobis' measures. As columns (3)-(4) of Table 11 show, greater inter-city technology spillovers are associated with increased patent similarity between connected cities. This finding suggests that inter-city technology spillovers induced by HSR connections shift the direction of innovation between connected cities toward greater similarity.

Taken together, our results show that HSR connections steer the innovation trajectories of connected city pairs toward greater similarity. Moreover, we provide suggestive evidence

[9]The sample mean of intercity patent similarity is 0.0676.

that technology spillovers facilitated by HSR connections contribute to this convergence in the direction of innovation.

8 Conclusion

Previous studies have demonstrated that advanced infrastructure promotes innovation by accelerating knowledge diffusion. However, most of these studies focus on overall innovations, typically measured by the quantity and quality of innovation outcomes. This study takes a different perspective by exploring how faster transportation affects the composition of local innovation.

We empirically investigate the effects of advanced transportation on the composition of local innovations, using the expansion of China's high-speed railway (HSR) as a natural experiment. Based on a large-scale dataset of Chinese patent filings from 2000 to 2017, we construct measures of important patents based on a large-language model, providing separate measures of incremental and breakthrough innovation. We find that HSR connections increase the share of incremental innovation while causing a reduction in the proportion of breakthrough innovation.

Furthermore, we explore the mechanisms behind the observed compositional shift between incremental and breakthrough innovation. We show that knowledge spillovers induced by HSR connections lead to asymmetric effects on incremental and breakthrough innovation, because increased knowledge spillovers reduce the relative cost in the exploitation of existing knowledge, compared to the exploration of unknown knowledge, incentivizing connected cities to undertake more incremental innovation. When we investigate how knowledge spillovers take place, we find evidence that knowledge spillovers are mainly driven by patent transactions, licensing, and collaborations.

In addition, we investigate how knowledge spillovers facilitated by HSR connections affect the technological direction of innovation between connected city pairs. We provide two pieces of evidence. First, we find that HSR connections lead to greater patent similarity between connected city pairs. Second, we show that HSR connections increase the knowledge spillovers between connected city pairs, which in turn significantly increase the between-city patent similarity. Together, these findings indicate that knowledge spillovers induced by HSR connectivity alter the trajectory of technological innovation in connected cities, steering it toward a more similar direction.

Overall, this study expands the understanding of how knowledge spillovers, induced by HSR connectivity, influence the relative composition between incremental and breakthrough

innovation. Our findings highlight the importance of HSR connectivity in shaping policy design aimed at promoting technology spillovers and innovation. Specifically, the existence of differential effects of HSR on incremental and breakthrough innovation has critical policy implications. For instance, the government should mitigate the negative effects of HSR connectivity to promote breakthrough innovations.

References

Abebe, Girum, Margaret McMillan, and Michel Serafinelli, "Foreign direct investment and knowledge diffusion in poor locations," *Journal of Development Economics*, 2022, *158*, 102926.

Abrams, David S, Ufuk Akcigit, and Jillian Popadak, "Patent value and citations: Creative destruction or defensive disruption," *NBER Working Paper Series*, 2013, (19763).

Acemoglu, Daron, Ufuk Akcigit, and Murat Alp Celik, "Radical and incremental innovation: The roles of firms, managers, and innovators," *American Economic Journal: Macroeconomics*, 2022, *14* (3), 199–249.

Agrawal, Ajay, Alberto Galasso, and Alexander Oettl, "Roads and innovation," *Review of Economics and Statistics*, 2017, *99* (3), 417–434.

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

Azoulay, Pierre, Waverly Ding, and Toby Stuart, "The determinants of faculty patenting behavior: Demographics or opportunities?," *Journal of Economic Behavior & Organization*, 2007, *63* (4), 599–623.

Bahar, Dany, Prithwiraj Choudhury, Do Yoon Kim, and Wesley W Koo, "Innovation on wings: Nonstop flights and firm innovation in the global context," *Management Science*, 2023, *69* (10), 6202–6223.

Bai, John, Wang Jin, and Sifan Zhou, "Proximity and knowledge spillovers: Evidence from the introduction of new airline routes," *Management Science*, 2024, *70* (11), 7464–7485.

Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang, "Roads, railroads, and decentralization of Chinese cities," *Review of Economics and Statistics*, 2017, *99* (3), 435–448.

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

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Boudreau, Kevin J, Tom Brady, Ina Ganguli, Patrick Gaule, Eva Guinan, Anthony Hollenberg, and Karim R Lakhani, "A field experiment on search costs and the formation of scientific collaborations," *Review of Economics and Statistics*, 2017, *99* (4), 565–576.

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson, "Product innovation, product diversification, and firm growth: Evidence from japan's early industrialization," *American Economic Review*, 2021, *111* (12), 3795–3826.

Byun, Seong K, Jong-Min Oh, and Han Xia, "Incremental vs. breakthrough innovation: The role of technology spillovers," *Management Science*, 2021, *67* (3), 1779–1802.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Capponi, Giovanna, Arianna Martinelli, and Alessandro Nuvolari, "Breakthrough innovations and where to find them," *Research Policy*, 2022, *51* (1), 104376.

Catalini, Christian, Christian Fons-Rosen, and Patrick Gaulé, "How do travel costs shape collaboration?," *Management Science*, 2020, *66* (8), 3340–3360.

Chen, Deqiu, Yujing Ma, Xiumin Martin, and Roni Michaely, "On the fast track: Information acquisition costs and information production," *Journal of Financial Economics*, 2022, *143* (2), 794–823.

Dong, Xiaofang, Siqi Zheng, and Matthew E Kahn, "The role of transportation speed in facilitating high skilled teamwork across cities," *Journal of Urban Economics*, 2020, *115*, 103212.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Hall, Bronwyn H and Rosemarie Ham Ziedonis, "The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995," *Rand Journal of Economics*, 2001, pp. 101–128.

Hanley, Douglas, Jiancheng Li, and Mingqin Wu, "High-speed railways and collaborative innovation," *Regional Science and Urban Economics*, 2022, *93*, 103717.

Jaffe, Adam B, "Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value," 1986.

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

Kaplan, Sarah and Keyvan Vakili, "The double-edged sword of recombination in breakthrough innovation," *Strategic Management Journal*, 2015, *36* (10), 1435–1457.

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

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring technological innovation over the long run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

Koh, Yumi, Jing Li, and Jianhuan Xu, "Subway, collaborative matching, and innovation," *Review of Economics and Statistics*, 2025, *107* (2), 476–493.

Lawrence, Martha, Richard Bullock, and Ziming Liu, *China's high-speed rail development*, World Bank Publications, 2019.

Lin, Yatang, "Travel costs and urban specialization patterns: Evidence from China's high speed railway system," *Journal of Urban Economics*, 2017, *98*, 98–123.

_ , **Yu Qin, and Zhuan Xie**, "Does foreign technology transfer spur domestic innovation? Evidence from the high-speed rail sector in China," *Journal of Comparative Economics*, 2021, *49* (1), 212–229.

Liu, Weiwei, Yuan Tao, and Kexin Bi, "Capturing information on global knowledge flows from patent transfers: An empirical study using USPTO patents," *Research Policy*, 2022, *51* (5), 104509.

Manso, Gustavo, "Motivating innovation," *The Journal of finance*, 2011, *66* (5), 1823–1860.

Maurseth, Per Botolf and Roger Svensson, "The importance of tacit knowledge: Dynamic inventor activity in the commercialization phase," *Research Policy*, 2020, *49* (7), 104012.

Nelson, Andrew J, "Measuring knowledge spillovers: What patents, licenses and publications reveal about innovation diffusion," *Research policy*, 2009, *38* (6), 994–1005.

Ollivier, Gerald, Richard Bullock, Jin Ying, and Nanyan Zhou, "High-speed railways in China," *China Transport Topics*, 2014, (11).

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

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Figure 1: China's High-Speed Rail (HSR) Network (2008–2020)

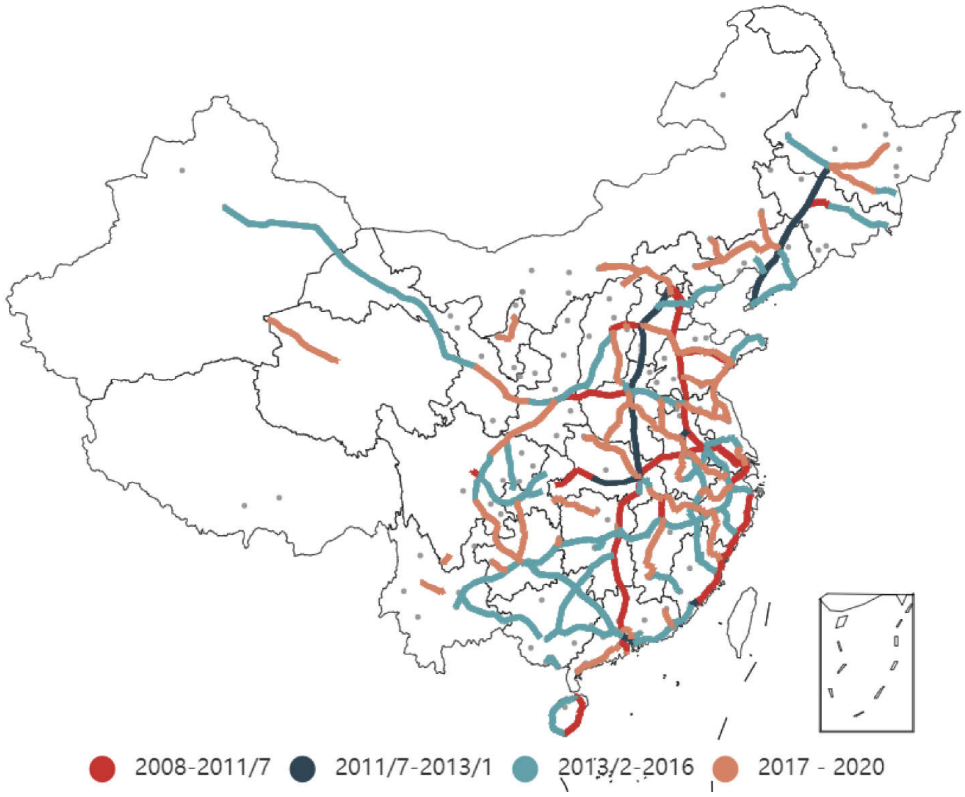
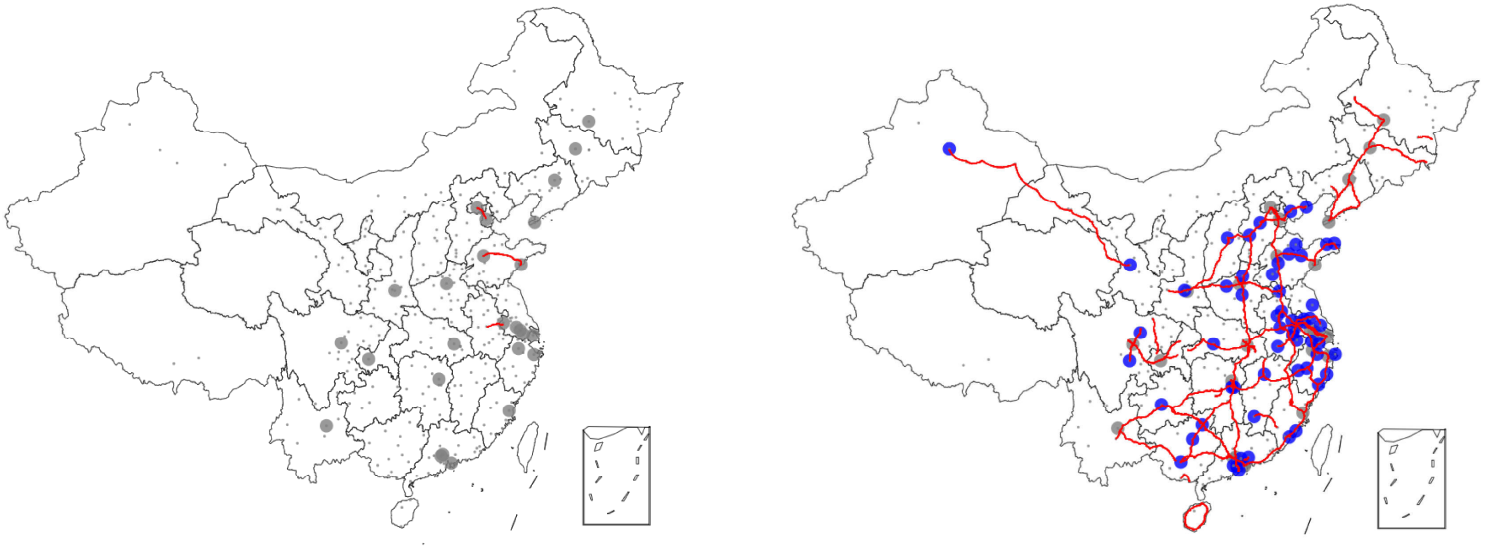


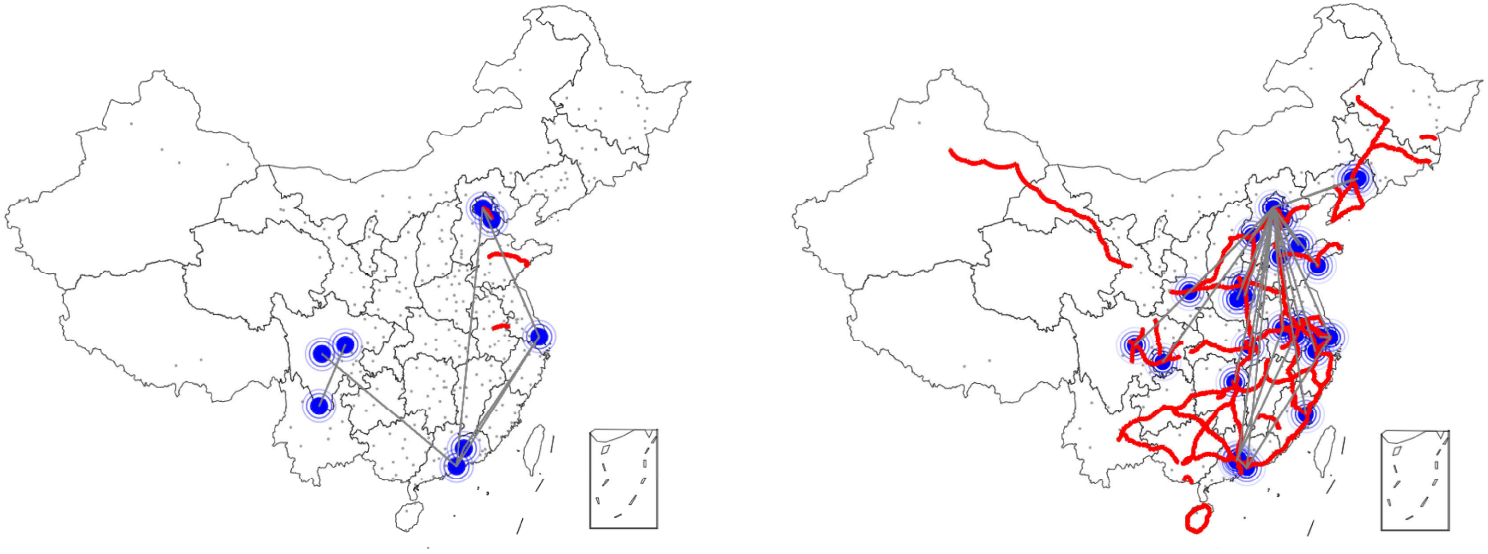
Figure 2: China's HSR and Granted Patent Filings (2008 and 2016)



(a) 2008 (36 cities)

(b) 2016 (56 cities)

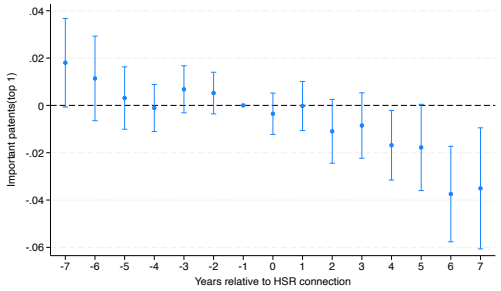
Figure 3: China's HSR and Cross-City Collaboration on Patents (2008 and 2016)



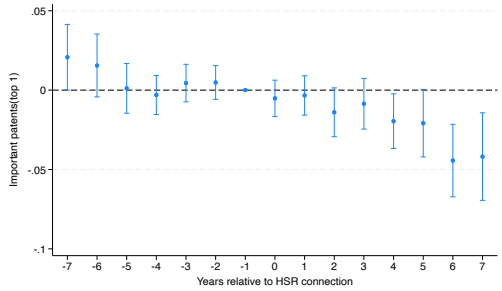
(a) 2008

(b) 2016

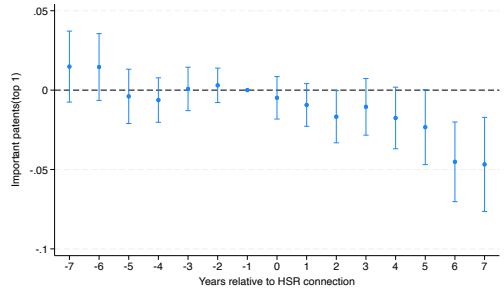
Figure 4: Event Study Graphics: Impact of HSR Connection on Incremental vs. Breakthrough Innovation



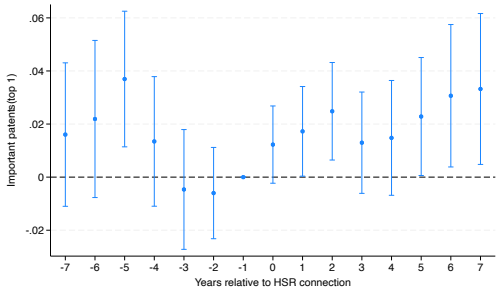
(a) Top Important Patents(15%)



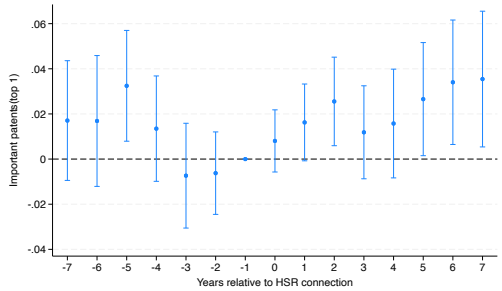
(b) Top Important Patents(20%)



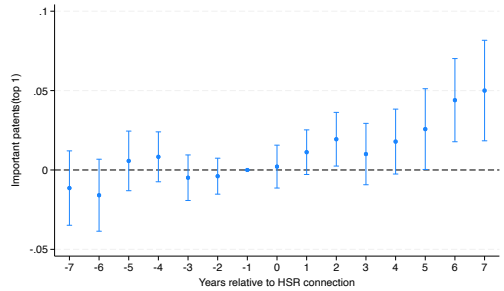
(c) Top Important Patents(25%)



(d) Bottom Important Patents(25%)

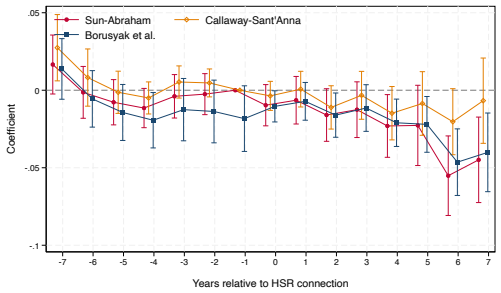


(e) Bottom Important Patents(30%)

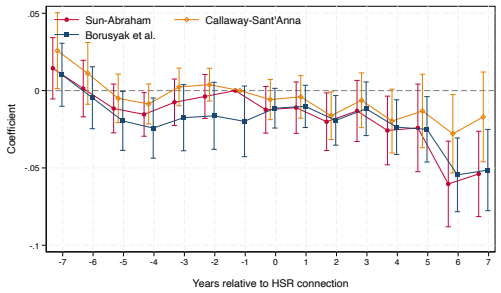


(f) Bottom Important Patents(70%)

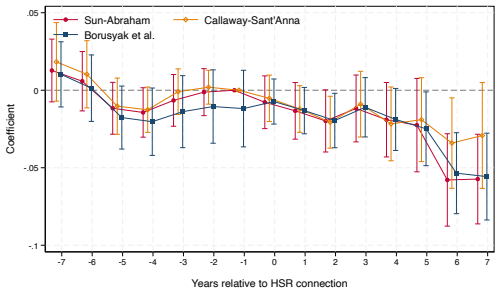
Figure 5: Event Study Graphics: using Robust DID Estimators



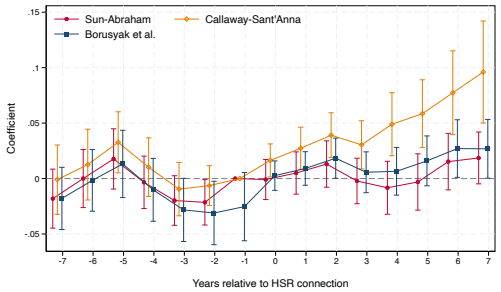
(a) Top Important Patents(15%)



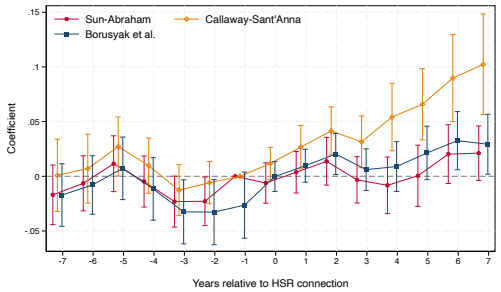
(b) Top Important Patents(20%)



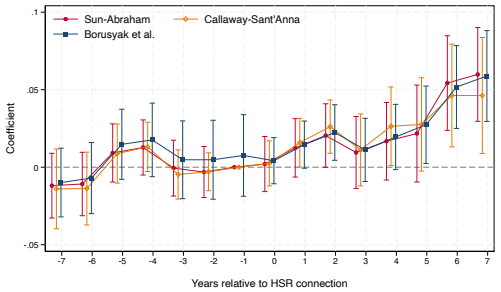
(c) Top Important Patents(25%)



(d) Bottom Important Patents(25%)

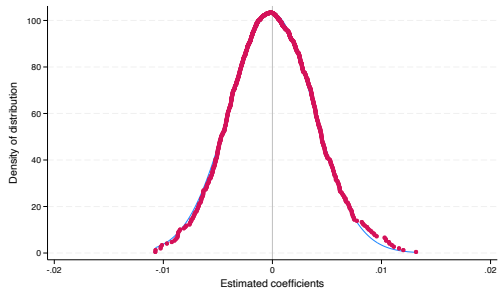


(e) Bottom Important Patents(30%)

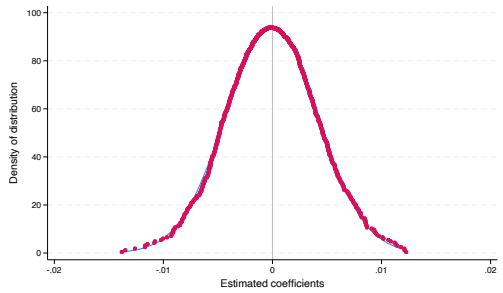


(f) Bottom Important Patents(70%)

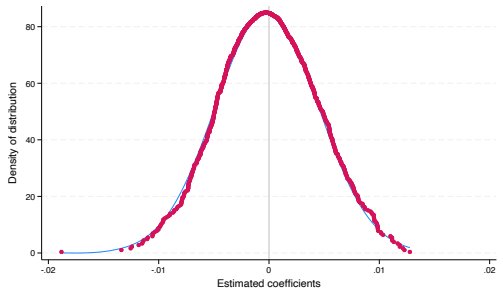
Figure 6: Distribution of Estimates in the Placebo Test (City Level)



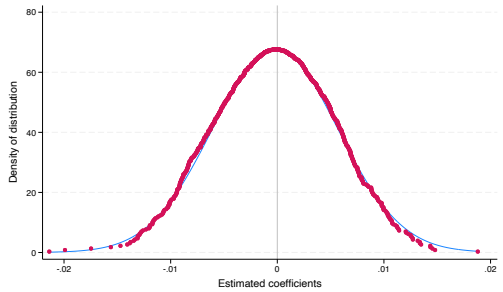
(a) Top Important Patents(15%)



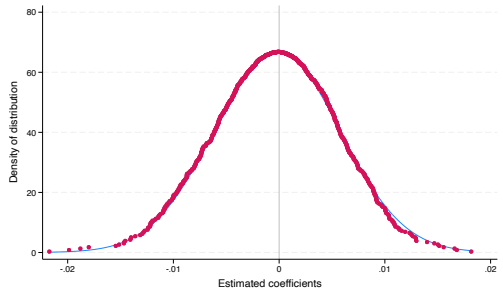
(b) Top Important Patents(20%)



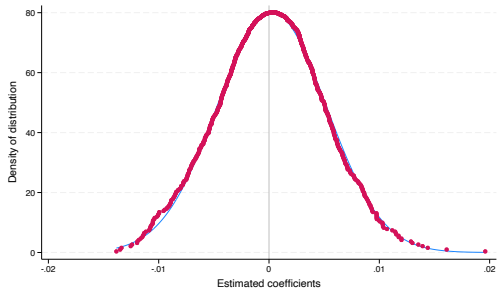
(c) Top Important Patents(25%)



(d) Bottom Important Patents(25%)



(e) Bottom Important Patents(30%)



(f) Bottom Important Patents(70%)

Table 1: Impact of HSR Connections on Incremental vs. Breakthrough Innovation

	Top Important Patents			Bottom Important Patents		
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 15%	Top 20%	Top 25%	Bottom 25%	Bottom 30%	Bottom 70%
L.HSR connection	-0.0126**	-0.0145**	-0.0162**	0.0140*	0.0161**	0.0185**
	(-2.32)	(-2.31)	(-2.36)	(1.96)	(2.18)	(2.57)
City-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4664	4664	4664	4664	4664	4664
R^2	0.360	0.405	0.434	0.602	0.594	0.465
Adj.R^2	0.315	0.363	0.394	0.575	0.565	0.428

t statistics in parentheses. Standard errors are clustered at city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Impact of HSR Connection on Local Innovation: Long Difference IV Regression

	Top Important Patents			Bottom Important Patents		
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 15%	Top 20%	Top 25%	Bottom 25%	Bottom 30%	Bottom 70%
IV = Historical Railroad						
HSR connection (instrumented)	-0.0387	-0.0911	-0.134*	0.105	0.103	0.161**
	(-0.74)	(-1.49)	(-1.93)	(1.34)	(1.24)	(2.14)
City-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	268	268	268	268	268	268
Kleibergen-Paap F-statistic	32.733	32.733	32.733	32.733	32.733	32.733

t statistics in parentheses. Standard errors are clustered at firm level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Robustness: Road and Air Ridership

	Top Important Patents			Bottom Important Patents		
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 15%	Top 20%	Top 25%	Bottom 25%	Bottom 30%	Bottom 70%
L.HSR connection	-0.0132**	-0.0152**	-0.0171**	0.0131*	0.0153**	0.0194***
	(-2.45)	(-2.45)	(-2.51)	(1.84)	(2.08)	(2.70)
Road ridership	0.000578	0.00176	0.00101	-0.000815	-0.00134	0.00160
	(0.18)	(0.50)	(0.26)	(-0.14)	(-0.25)	(0.37)
Air ridership	-0.00112	-0.00136	-0.00230	-0.000723	-0.000237	0.00210
	(-0.79)	(-0.83)	(-1.33)	(-0.39)	(-0.12)	(1.11)
City-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4642	4642	4642	4642	4642	4642
R^2	0.361	0.406	0.435	0.602	0.595	0.466
Adj.R^2	0.316	0.364	0.395	0.574	0.566	0.428

t statistics in parentheses. Standard errors are clustered at city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Robustness: ICT development

	Top Important Patents			Bottom Important Patents		
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 15%	Top 20%	Top 25%	Bottom 25%	Bottom 30%	Bottom 70%
L.HSR connection	-0.0118**	-0.0133**	-0.0150**	0.0141*	0.0164**	0.0172**
	(-2.14)	(-2.12)	(-2.18)	(1.94)	(2.18)	(2.38)
Broadband subscriber	0.00244	0.00192	-0.000287	0.000907	0.00479	0.000312
	(0.44)	(0.30)	(-0.04)	(0.12)	(0.66)	(0.04)
City-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4593	4593	4593	4593	4593	4593
R^2	0.358	0.404	0.433	0.601	0.592	0.464
Adj.R^2	0.312	0.362	0.393	0.573	0.564	0.426

t statistics in parentheses. Standard errors are clustered at city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The Impact of HSR connection on Technology Spillovers

	Technology spillover(5%)		Technology spillover(15%)	
	(1)	(2)	(3)	(4)
	Technology spillover(Jaffe)	Technology spillover(Mahalanobis)	Technology spillover(Jaffe)	Technology spillover(Mahalanobis)
L.HSR connection	0.289***	0.275***	0.303***	0.295***
	(5.22)	(4.97)	(5.18)	(4.99)
City-level Controls	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	4664	4664	4664	4664
R^2	0.975	0.974	0.970	0.969
Adj.R^2	0.973	0.972	0.968	0.967

t statistics in parentheses. Standard errors are clustered at city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: The Role of Technology Spillovers in Shaping Innovation

	Top Important Patents			Bottom Important Patents		
	(1) Top 15%	(2) Top 20%	(3) Top 25%	(4) Bottom 25%	(5) Bottom 30%	(6) Bottom 70%
Panel A: Jaffe distance						
L.SpillTech(Jaffe distance)	-0.0166***	-0.0172***	-0.0191***	0.00276	0.00478	0.0184***
	(-3.55)	(-3.42)	(-3.22)	(0.48)	(0.82)	(2.80)
City-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4664	4664	4664	4664	4664	4664
R^2	0.335	0.363	0.437	0.602	0.594	0.467
Adj.R^2	0.289	0.319	0.397	0.574	0.565	0.430
Panel B: Mahalanobis distance						
L.SpillTech(Mahalanobis Distance)	-0.0153***	-0.0171***	-0.0177***	0.00351	0.00560	0.0169**
	(-3.04)	(-3.03)	(-2.94)	(0.61)	(0.97)	(2.55)
City-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4664	4664	4664	4664	4664	4664
R^2	0.363	0.408	0.436	0.602	0.594	0.467
Adj.R^2	0.318	0.366	0.397	0.574	0.565	0.429

t statistics in parentheses. Standard errors are clustered at city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Inter-city Patent Citations, Transfers, and Licensing (city-pair level)

	Inter-city citations		Patent transfer	Patent licensing
	(1)	(2)	(3)	(4)
	Inter-city citations	Inter-city citations of university patents	Patent transfers	Patent licensing
L.HSR connection	0.198***	0.144***	0.0636***	0.0118***
	(44.79)	(36.56)	(25.39)	(9.81)
City-pair Controls	Yes	Yes	Yes	Yes
City-pair FE	Yes	Yes	Yes	Yes
Citing city*Year FE	Yes	Yes	Yes	Yes
Cited city*Year FE	Yes	Yes	Yes	Yes
Observations	1317968	1317968	1317968	1317968
R^2	0.805	0.779	0.459	0.307
Adj.R^2	0.790	0.763	0.420	0.256

t statistics in parentheses. Standard errors are clustered at city-pair level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Inter-city Patent Collaborations (city-pair level)

	Collaborations		New Collaborations	
	(1)	(2)	(3)	(4)
	Collab. Patents	Collab. Teams	New Collab. Teams	Patents by New Teams
L.HSR connection	0.0719***	0.0527***	0.0446***	0.0601***
	(15.34)	(16.28)	(15.97)	(15.31)
City-pair Controls	Yes	Yes	Yes	Yes
City-pair FE	Yes	Yes	Yes	Yes
Citing city*Year FE	Yes	Yes	Yes	Yes
Cited city*Year FE	Yes	Yes	Yes	Yes
Observations	658832	658832	658832	658832
R^2	0.635	0.647	0.596	0.553
Adj.R^2	0.605	0.618	0.563	0.516

t statistics in parentheses. Standard errors are clustered at city-pair level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Effects of HSR Connections on Direction of Innovation across Cities

	Direction of Technological Innovaton	
	(1) Similarity	(2) Similarity
L.HSR connection	0.0123***	0.0121***
	(33.21)	(33.26)
City-pair Controls	No	Yes
City-pair FE	Yes	Yes
Citing-city*Year FE	Yes	Yes
Cited-city*Year FE	Yes	Yes
Observations	663080	658832
R^2	0.968	0.968
Adj.R^2	0.966	0.966

t statistics in parentheses. Standard errors are clustered at city-pair level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Long Difference IV Regressions: Inter-city Similarity

	Long-differenced IV	First-stage
	(1)	(2)
	Inter-city similarity	HSR connection
HSR connection	0.0519***	
	(11.09)	
Historical Railroad		0.292***
		(17.49)
City-pair Controls	Yes	Yes
Province-pair FE	Yes	Yes
Observations	37942	37942
R^2	0.062	0.249
Kleibergen-Paap F-statistic	305.780	

t statistics in parentheses. Standard errors are clustered at province-pair level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Inter-city Technology Spillovers and Direction of Innovation across Cities

	Inter-city technology spillovers		Impact of Inter-city Technology Spillovers	
	(1)	(2)	(3)	(4)
	Jaffe Measure	Mahalanobis Measure	Inter-city Similarity	Inter-city Similarity
L.HSR connection	0.0182***	0.0110***		
	(6.59)	(5.95)		
L.Intercity technology spillovers(Jaffe measure)			0.00857***	
			(25.58)	
L.Intercity technology spillovers(Mahalanobis measure)				0.0111***
				(20.16)
City-pair Controls	Yes	Yes	Yes	Yes
City-pair FE	Yes	Yes	Yes	Yes
Citing city*Year FE	Yes	Yes	Yes	Yes
Cited city*Year FE	Yes	Yes	Yes	Yes
Observations	658832	658832	658832	658832
R^2	0.998	0.999	0.968	0.968
Adj.R^2	0.997	0.999	0.966	0.966

t statistics in parentheses. Standard errors are clustered at city-pair level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix

A Mahalanobis Measure of Technology Spillovers OA-1

B A Conceptual Model OA-2

A Mahalanobis Measure of Technology Spillovers

The Mahalanobis measure of technology spillovers is based on the Mahalanobis distance of technological proximity, as proposed by Bloom et al. (2013).

To begin with, we construct the technological proximity based on Mahalanobis distance following Bloom et al. (2013). This approach offers a more comprehensive assessment of technological relatedness by capturing knowledge spillovers both within and across technology classes. The detailed procedure is as follows:

In the first step, we define a $T \times N$ matrix U_t in year t.

$$U_t = [U_1^{'}, ...U_i^{'}, ..., U_N^{'}]$$

and we also define a normalized $T \times N$ matrix $\tilde{U}_t$ in year t.

$$\tilde{U}_t = [\frac{U_1^{'}}{U_1 U_1^{'}}, ..., \frac{U_i^{'}}{U_i U_i^{'}}, ..., \frac{U_N^{'}}{U_N U_N^{'}}]$$

where $U_{i,t} = (u_{i1,t}, u_{i2,t}, ..., u_{i\tau,t}, ..., u_{iT,t})$ is a vector that denotes city i's proportion of patents in each technology class τ at 3-digit IPC level in year t. For notational simplicity, we simply denote $U_{i,t}$ as U_i. And, each column is normalized by the dot product of each city's patent share. Here, T denotes the total number of technology classes, and N denotes the total number of cities in our sample.

In the second step, we define a $N \times N$ matrix $TECH = \tilde{U}_t^{'} \tilde{U}_t$, which is the standard Jaffe uncentered correlation measure between city i and city j.

In the third step, we define a $N \times T$ matrix $\tilde{X}_t = [\frac{U_{(:,1)}^{'}}{(U_{(:,1)} U_{(:,1)}^{'})^{1/2}}, ..., \frac{U_{(:,T)}^{'}}{(U_{(:,T)} U_{(:,T)}^{'})^{1/2}}]$, where $U_{(:,i)}$ is a $1 \times N$ vector, which is the ith row of U_t. $\tilde{X}_t$ is the normalized patent class share across cities rather than city share across patent classes.

Forth, we define a $T \times T$ matrix $\Omega_t = \tilde{X}^{'} \tilde{X}$, in which each element is the standard Jaffe uncentered correlation measure between patent classes. Ω_t captures technology spillovers both within and across technology classes.

In the fifth step, we obtain the technological proximity based on Mahalanobis distance between city i and city j, which is specified as follows:

$$TECH_t^{Maha} = \tilde{U}_t^{'} \Omega_t \tilde{U}_t \tag{9}$$

where $TECH^{Maha}$ is an $N \times N$ matrix, and the (i, j) element, denoted as $tech_{ijt}^{Maha}$, measures

the correlation weights of the overlaps in patent share between city i and city j and captures how close their patent technology classes are to each other. In particular, if $\Omega = I$, then $TECH_t^{Maha} = TECH_t^{Jaffe}$.

Next, we construct a modified measure of technology spillovers, based on the Mahalanobis distance of technological proximity, as proposed by Bloom et al. (2013). Specifically, Mahalanobis measure of technology spillovers from all other cities to city i is specified as follows:

$$SpillTech_{i,t}^{Maha} = \sum_{i \neq j} tech_{ij,t}^{Maha} Stock_{j,t} \tag{10}$$

where $tech_{ij,t}^{Maha}$ is the technological proximity based on the Mahalanobis distance between city i and city j in year t. $Stock_{j,t}$ is the stock of knowledge in city j, which is calculated with a perpetual inventory method with a $\delta = 15\%$ depreciation rate. A higher value of $SpillTech_{i,t}^{Maha}$ indicates a greater amount of technology spillovers, which captures spillovers both within and across technology classes.

B A Conceptual Model

We construct a conceptual framework to illustrate the different effects of infrastructure on two types of innovations, providing a framework for empirical analysis in the following section.

Environment. Consider two distant cities ($i = 1, 2$), each with a representative firm that allocates resources between incremental and breakthrough innovations. The knowledge production functions exhibit constant returns to scale, as specified by Equation (11).

$$\begin{aligned} \text{Incremental:} \quad & y_i = \ell^{1-\phi_i} [m^{\gamma} n^{1-\gamma}]^{\phi_i} \\ \text{Breakthrough:} \quad & y_b = \ell^{1-\phi_b} [m^{\gamma} n^{1-\gamma}]^{\phi_b} \end{aligned} \tag{11}$$

The firm in each city combines distinct knowledge m from City 1 and n from City 2, along with other essential inputs ℓ such as scientists, capital, and laboratories, to produce innovation outputs y_i and y_b. Here, ϕ_i and ϕ_b represent the knowledge intensity for the two types of innovations. Following Manso (2011), we assume incremental innovation is more intensive in existing knowledge ($\phi_i > \phi_b$). The parameter γ represents the share of local knowledge.

Firms' problem. Firms in both cities maximize their expected profits as specified by the following equations:

$$\begin{aligned}\pi_1 =& \chi \ell_{1i}^{1-\phi_i}[m_{1i}^{\gamma} n_{1i}^{1-\gamma}]^{\phi_i} - w_1 \ell_{1i} + \ell_{1b}^{1-\phi_b}[m_{1b}^{\gamma} n_{1b}^{1-\gamma}]^{\phi_b} - w_1 \ell_{1b} \\ \pi_2 =& \chi \ell_{2i}^{1-\phi_i}[n_{2i}^{\gamma} m_{2i}^{1-\gamma}]^{\phi_i} - w_2 \ell_{2i} + \ell_{2b}^{1-\phi_b}[n_{2b}^{\gamma} m_{2b}^{1-\gamma}]^{\phi_b} - w_2 \ell_{2b}\end{aligned} \tag{12}$$

The final output of breakthrough innovations is multiplied by $\chi > 1$, as breakthrough innovation yields higher profits than incremental innovation. The cost of innovative factors (ℓ), such as scientific funds and qualified scientists, is denoted by w. Knowledge stocks M and N in Cities 1 and 2, respectively, are public goods freely available to firms. Consequently, firms will fully utilize local knowledge.

$$m_{1i} = m_{1b} = M, \quad n_{2i} = n_{2b} = N \tag{13}$$

However, firms can only acquire a fraction ($\rho < 1$) of the knowledge from the distant city due to geographic barriers that hinder knowledge diffusion.

$$n_{1i} = n_{1b} = \rho N, \quad m_{2i} = m_{2b} = \rho M \tag{14}$$

The firm faces a constraint on total resources available locally for innovation ($\overline{L}$), since it is difficult for scientific funds, laboratories, or qualified scientists to move across locations.

$$\ell_{1i} + \ell_{1b} = \overline{L}_1, \quad \ell_{2i} + \ell_{2b} = \overline{L}_2 \tag{15}$$

Local innovations. We consider two measurements to reflect major aspects of local innovation in our model. First, we use the ratio of incremental to breakthrough innovations to represent the composition of local innovations:

$$R_1 = \frac{y_{1i}}{y_{1b}}, \quad R_2 = \frac{y_{2i}}{y_{2b}} \tag{16}$$

Second, we define the similarity in innovation between Cities 1 and 2 based on input mix, following Boehm et al. (2022):

$$CosSim(\vec{v}_1, \vec{v}_2) = \frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\| \cdot \|\vec{v}_2\|}, \quad \text{where } \vec{v}_c = \begin{pmatrix} \ell_c \\ m_c \\ n_c \end{pmatrix}, \quad c \in \{1, 2\} \tag{17}$$

Let $\vec{v}_c$ represent a bundle of factors, including innovative resources (ℓ_c) and existing knowledge from Cities 1 and 2 (m_c and n_c). We use the cosine similarity between vectors for the two cities to measure the inter-city similarity of innovation output.

Based on the model setting, we find a positive relationship between a firm's access to knowledge in the distant city and the ratio of incremental to breakthrough innovations. Specifically, R_1 and R_2 increase with the fraction of available distant knowledge (ρ). This is because incremental innovation, being knowledge-intensive, benefits more from enhanced access to existing knowledge compared to breakthrough innovation. This finding is summarized by the following proposition:

Proposition 1. Advanced transportation benefits incremental innovations

Advanced transportation infrastructure, such as the introduction of high-speed rails, enhances firms' access to existing knowledge in distant cities. Consequently, the proportion of incremental innovations increases in both cities.

$$\rho \uparrow \implies R_1 \uparrow, R_2 \uparrow$$

By substituting Equations (3)–(5) into (7), we can demonstrate with our model that increased access to distant knowledge (ρ) leads to a higher cosine similarity between the input bundles in two cities, indicating greater inter-city similarity in innovation. Faster transportation enables firms in distant locations to share a similar mix of factors, leading to increased similarity in innovation outcomes across cities. This result is summarized by the following proposition:

Proposition 2. Advanced transportation increases inter-city similarity

Knowledge diffusion facilitated by advanced transportation infrastructure aligns the composition of factors used for innovation, resulting in greater similarity between innovations in two cities.

$$\rho \uparrow \implies CosSim(\vec{v}_1, \vec{v}_2) \uparrow$$

Therefore, in this section, we examine whether knowledge spillovers induced by HSR connections affect the direction of technological innovation, as measured by inter-city patent similarity.