Knowledge Diffusion and Local Innovation: Evidence from China

Yousha Liang Kang Shi Junjie Tang Hanyi Tao Juanyi Xu*

September, 2025

Abstract

This study examines how the high-speed rail (HSR) network shapes the focus of local innovation by accelerating technology spillovers. By merging large-scale textual analyses of Chinese patents with quasi-experimental variations stemming from China's HSR expansion, we show that improved inter-city rail links shift inventive activities toward incremental innovation rather than breakthrough advances. We attribute this shift to technology spillovers enabled by HSR, which lower the cost of incremental innovation relative to breakthrough innovation and, consequently, alter the innovation mix. Patent transactions, licensing, and collaborations emerge as the key channels transmitting these spillovers across cities. As a result, HSR-connected cities produce increasingly similar inventions, revealing a new geography of innovation shaped by fast rail corridors.

JEL Code: O31, O33, O18.

Keywords: Breakthrough 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

Transportation infrastructure is widely believed to promote regional innovation and economic growth (Duranton and Turner, 2012; Agrawal et al., 2017). While the benefits of transportation infrastructure are typically evident, its unintended consequences on innovation are less well understood. In particular, innovation varies in its reliance on existing knowledge (Manso, 2011). *Incremental innovation* involves exploiting existing knowledge to make improvements within a particular "technology cluster". While such progress boosts productivity with lower risks, it eventually leads to 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 innovation. Moreover, breakthrough innovation is vital to long-term economic growth, as it enhances production efficiency and inspires future innovation (Acemoglu et al., 2022; Abebe et al., 2022). Easier access to existing knowledge induced by improved transportation infrastructure may lower the cost of exploiting known knowledge relative to exploring unknown areas, thereby shifting the composition of incremental versus breakthrough innovation (Akcigit and Kerr, 2018). Therefore, this paper aims to examine how the knowledge diffusion induced by HSR connections–one of the ambitious transportation networks in China–shapes the composition of innovation.

To this end, we use the rapid expansion of China's high-speed rail (HSR) network as a quasi-natural experiment to examine its impact on the composition of local innovation. We construct a matched dataset that includes information on HSR constructions and patent filings from 2000 to 2017. Our dataset covers 2.2 million Chinese patents granted in 285 Chinese prefecture-level cities, coupled with the introduction of 95 HSR lines across cities by the end of 2017.[1]

To assess the impact of HSR connections on the composition of local innovation, we propose the importance index to gauge the degree of breakthrough in local innovation, as suggested by Kelly et al. (2021). To begin with, based on the text of patent abstracts and using textual analysis techniques, we estimate the patent-level importance index by comparing each patent's distinction from its predecessors and its similarity to its successors. The rationale is that breakthrough innovation tends to depart from previous inventions and leads to various follow-up inventions. 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 1024-dimensional vector for the input text. We measure a patent's similarity to

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

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 top and bottom important patents for each city and each year. These shares are used to measure the composition of breakthrough and incremental innovation, respectively.

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 innovation composition 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-difference regressions, to further mitigate these concerns and enhance the robustness of our estimates.

We find that more connectivity induced through HSR operations disproportionately promotes incremental innovation over breakthrough innovation, resulting in a change in the composition of local innovation. 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 their respective sample means. Furthermore, the results of the event study specification show that, before the introduction of HSR, the proportions of both top and bottom important patents follow similar trends in HSR-connected cities and unconnected cities. This finding supports the validity of the identifying assumption. We also find that the impact of HSR connectivity on the share of breakthrough innovation is initially neutral or mildly negative, but progressively becomes more detrimental over time. In contrast, the positive effect of HSR connectivity on the share of incremental innovation is

initially modest but grows over time.

Our findings remain unchanged 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 long-difference regressions, 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 support this conclusion. First, we find that HSR connections enhance technology spillovers, as measured by both the Jaffe measure and the 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 technology spillovers across cities induced by HSR connections.

Having established that technology spillovers help explain the changes in the composition of local innovation, we further explore how technology spillovers occur between cities. Our analyses at the city-pair level show that HSR connections lead to more inter-city 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 between cities.

Furthermore, we explore whether technology spillovers induced by HSR connections affect the direction of technological innovation, as measured by inter-city 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 technology spillovers induced by HSR connections change the trajectory of innovation, promoting greater similarity in technological development.

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 output, 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 view in this literature is that knowledge diffusion increases both innovation quantity and quality. Our analysis extends this line of inquiry by revealing an understudied effect: while knowledge diffusion elevates total innovation, it disproportionately incentivizes incremental innovation relative to breakthrough innovation. By identifying this compositional shift, our study highlights the role of technology spillovers in shaping the *composition* of local innovation, rather than merely its scale. Additionally, we explore how technology spillovers take 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 the choices between incremental and breakthrough innovation. Earlier studies have concentrated on firm-level attributes and incentive schemes. For example, Manso (2011) and Acemoglu et al. (2022) examine how long-term compensation and job security incentivize researchers to pursue riskier, radical innovation. These models share a key premise: breakthrough ideas draw less heavily on the current stock of knowledge than incremental ones. We extend this literature in two directions. First, we identify an overlooked determinant of the incremental-to-breakthrough mix: improvements in transportation infrastructure. By lowering the cost of accessing distant knowledge, better transport allocates more innovative efforts toward incremental projects. Second, we provide empirical support for the key assumption used by Manso (2011) and Acemoglu et al. (2022): technology spillovers discourage exploration into the unknown (breakthroughs) while reinforcing exploitation of existing knowledge (incrementals). These findings strengthen the micro-mechanisms that may underlie observed shifts in the composition of local innovation.

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 networks increases academic publications in China. Koh et al. (2025) find that rapid expansion of the subway in Beijing boosts patent collaborations. Although prior research has documented that better transportation infrastructure boosts overall local innovation, it has not examined how improved connectivity changes the composition of that innovation. We provide the first evidence on this question. Leveraging a large-scale textual analysis of Chinese patent data, we classify each invention as either incremental or breakthrough. We demonstrate that HSR connections significantly alter the composition of local innovation: the share of breakthrough inventions falls while that of incremental inventions rises.

Finally, our study advances the literature that develops rigorous methods for identifying technological breakthroughs. The existing literature has yet to reach a consensus on the identification of breakthrough innovation. Earlier work has relied primarily on patent citations to gauge a patent's influence. For instance, forward citations, as well as measures of originality and generality, have been widely used in the literature (Hall et al., 2001; Lerner and Seru, 2022; Acemoglu et al., 2022). However, patent citation data are subject to well-documented biases, including home bias (Jaffe et al., 1993; Thompson and Fox-Kean, 2005; Griffith et al., 2011), random citation (Jaffe et al., 2000), and strategic citation (Lampe, 2012), which can distort inference. Consequently, measures that rely on patent citations may yield biased conclusions. In response, recent studies have begun to leverage textual analysis to identify breakthrough innovation (Kelly et al., 2021). Specifically, Kelly et al. (2021) quantify the extent of breakthrough for patents using term frequency–inverse document frequency (TF-IDF) algorithms, which assess the textual distinction of a patent from prior and subsequent work. Our study builds on this methodology by applying Large Language Models (LLMs) to patent text analysis. While Kelly et al. (2021) measure lexical divergence based on the frequency of words, we employ LLMs to evaluate semantic originality, capturing the contextual meaning of the text rather than merely assessing word frequency. For example, the TF-IDF method would treat "dog bites human" and "human bites dog" as identical due to exactly the same shared terms, whereas LLMs recognize their distinct meanings. This shift from syntactic to semantic analysis enhances precision in identifying breakthrough innovation.

The remainder of the paper is organized as follows. Section 2 provides essential background on China's HSR development. Section 3 describes the data sources and defines the key variables. 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) network has undergone unprecedented expansion since 2007, reaching 48,000 kilometers by 2024—representing two-thirds of global HSR infrastructure. This transformation creates a compelling quasi-experimental setting to study how HSR connectivity affects local innovation through enhanced knowledge spillovers. Three features about China's HSR enable our study: (i) rapid HSR network expansion, (ii) centralized planning process ensuring plausibly exogenous HSR construction, and (3) graphical evidence linking HSR to innovation and knowledge spillover patterns.

Rapid HSR Expansion in China. China's HSR has developed rapidly since 2007, driven by strong top-down policy support and efficient implementation by local governments. The HSR network experienced a remarkable expansion during 2008-2017, as illustrated by Figure 1. This figure shows the geographic development of China's HSR network during this period. In particular, by December 2012, the two most important lines, the Beijing-Shanghai line and the Beijing-Guangzhou line, were completed. These lines connect the three most dynamic economic clusters in China. By the end of 2017, the HSR network had reached 25,000 km, which accounts for 66 percent of the world's total HSR lines (Lawrence et al., 2019). Moreover, during the decade, the operational length of the HSR tracks grew at an average annual rate of 16.9%, while HSR ridership surged at a striking compound annual growth rate of 33.5%. By 2024, HSR had become the primary mode of passenger transport in China, accommodating approximately 3.3 billion passengers and accounting for over 60% of China's total rail ridership.

The unprecedented scale and speed of China's HSR development create an ideal natural experiment to investigate the impact of improved transportation infrastructure on local innovation.

Centralized Planning in HSR Development. The design of China's HSR routes was primarily determined by the Medium- and Long-Term Railway Plan (MLTRP), introduced by the central government in 2004. The plan aimed to establish a four-horizontal-four-vertical HSR grid connecting all major cities. It reflected strategic national objectives, including

aggregate growth, regional balance, and security, with implementation characterized by a high degree of centralization.

This institutional structure sharply curtailed local discretion (Lawrence et al., 2019). Local governments were required to align infrastructure investments with the MLTRP, while deviations from approved routes were systematically discouraged to preserve national uniformity. Moreover, construction proceeded according to a rigid multi-year timetable, largely insulated from contemporaneous local economic conditions, including innovation activity.

This combination of centralized planning and implementation rigidity provides a quasi-experimental setting in which HSR expansion can be plausibly treated as independent of city-level economic shocks. As a result, the staggered expansion of China's HSR network enables us to study the impact of HSR connections, which are likely to be accidental outcomes, on local innovation (Lin et al., 2023).

HSR and Local Innovation Activities China's rapid expansion of the HSR coincided with a notable acceleration in innovative activities. Between 2008 and 2017, parallel to the peak phase of HSR network development, city-level patent filings grew at a compound annual rate of 15%, based on our comprehensive patent dataset.

Figure 2 provides suggestive spatial evidence of this relationship, comparing the distribution of cities with above-average patent filings before and after the expansion of major HSR routes.[2] The comparison between 2008 and 2017 conveys one important message that although the 20 newly emerged highly innovative cities (blue dots in Figure 2(b)) are geographically dispersed, they are closely aligned with newly constructed HSR routes (red lines). This spatial correlation motivates our investigation into whether and how improved HSR connectivity shapes local innovation.

Additional graphical evidence suggests a potential mechanism: the effects of HSR may be realized through increased inter-city knowledge spillovers. Figure 3 displays the geographic distribution of cities with above-average inter-city patent citations in 2008 and 2017, respectively. In particular, Figure 3(b) reveals a marked overlap between the expansion of the HSR network (red lines) and inter-city patent citations (gray lines), with citation density visibly higher among HSR-connected city pairs. In addition, Figure 4 displays the geographic distribution of cities with above-average inter-city collaborative patents in 2008 and 2017, respectively. In particular, Figure 4(b) demonstrates a similar positive relationship between

[2]Here, cities with above-average patent filings are those whose patent filings exceed the sample mean of 550 patents.

HSR connections (red lines) and inter-city patent collaborations (gray lines).

While these patterns do not establish causality, they provide a conceptual roadmap for our further formal analysis. In particular, we subsequently test whether HSR-induced technology spillovers contribute to compositional change in local innovation in China in Section 6.

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 2000 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 applicants' bibliography information, filing date, grant date, abstract, 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 employing Large Language Models to analyze the patent texts, we construct a measure of patent importance index, which is 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 2004, 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 2000 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 the Chinese administrative hierarchy, they typically govern larger rural areas, while prefecture-level cities are more industrialized. As a result, 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 Chinese granted

patents in 285 Chinese prefecture-level cities from 2000 to 2017. These cities are connected by 95 HSR lines by the end of 2017. We choose 2017 as the cutoff year to enable a comprehensive evaluation of each patent's influence by comparing it with its successors over the following 3 years.

3.2 Key Variables

HSR Connections. Using the operation time of HSR stations and HSR lines, we construct the city-level indicator of HSR connections for cities at the prefecture level. First, based on the operation time of each HSR line, we begin by identifying the earliest year that a city i is connected by the HSR lines. We then construct the city-level dummy variable for HSR connections, $HSR_{i,t}$, which takes a value of 1 if city i is connected to any HSR line in year t and 0 otherwise.

Patent Importance Index. In the baseline analysis, we construct an importance index for each patent following the approach proposed by Kelly et al. (2021). We first collect the text of each patent abstract and obtain the embedding of each piece of text using the Large Language Model *embedding-2* developed by Zhipu AI, which generates a 1024-dimensional vector to represent the input text in the vector space. Then, we calculate the patent importance index for each patent by comparing its semantic similarity with both its predecessors and its successors, as defined in Equation (1).

$$\text{Patent Importance Index}_{p,t} = \frac{\frac{1}{N_g} \sum_{g \in x, s \in [t+1, t+3]} cosine(\vec{\nu}_{p,t}, \vec{\nu}_{g,s})}{\frac{1}{N_q} \sum_{q \in x, s \in [t-3, t-1]} cosine(\vec{\nu}_{p,t}, \vec{\nu}_{q,s})} \tag{1}$$

where $cosine(\vec{\nu}_{p,t}, \vec{\nu}_{q,s})$ represents the cosine similarity between patent p filed in year t and patent q filed in year s. The patent vectors $\vec{\nu}_{p,t}$ and $\vec{\nu}_{q,s}$ are constructed using LLMs, based on the patent texts of patent p and patent q, respectively. Specifically, for each patent p filed in year t, we calculate its average cosine similarity with patents q that share the same technology class and are filed between years $t-3$ and $t-1$, reflecting the extent to which patent p builds on existing knowledge. Similarly, we also calculate the average cosine similarity between patent p and patents g that share the same technology class and are filed between years $t+1$ and $t+3$, which is denoted as $cosine(\vec{\nu}_{p,t}, \vec{\nu}_{g,s})$. This metric captures the degree to which patent p explores uncharted areas.

The patent importance index identifies patents that are substantially novel relative to

their predecessors and significantly influence subsequent innovation. A higher value of the importance index signifies patents that are both novel and influential. Specifically, novel and influential patents that show high similarity to their successors but low similarity to their predecessors, indicating that they build on less existing knowledge while exploring more unknown areas, are regarded as breakthrough innovation. Therefore, the patent importance index effectively identifies breakthrough patents characterized by substantial novelty and significant influence on technological advancements.

A notable difference of our study from Kelly et al. (2021) is that we enhance their 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. The LLM offers several advantages over TF-IDF in the context of text embedding. First, the LLM captures semantic meaning within the text, whereas TF-IDF is based solely on term frequency and does not account for contextual meaning. Second, 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. In the baseline analysis, we adopt the embedding model *embedding-2* developed by Zhipu AI, which generates a 1024-dimensional vector to represent the input text in the vector space.[3]

Next, we calculate the share of important patents for each city. Specifically, we first identify patents that are above the Q^{th} percentile of the distribution based on *patent important index*. Second, we calculate city i's share of important innovation for each year as defined in Equation (2).

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

where $z_{it}(Q)$ represents the number of city i's patents that are above the Q^{th} percentile of the important index. Intuitively, a higher *Top Important Share* indicates that city i undertakes more *important patents*, reflecting the share of breakthrough innovation. In the baseline analysis, we classify the top 15 percent, 20 percent, and 25 percent of patents 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 Spillovers. We measure the magnitude of technology spillovers across cities based on the methodology proposed by Jaffe (1986) and Bloom et al. (2013). The magnitude of technology spillovers from city j to city i depends on two aspects. First, a key component

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

is the level of knowledge stock in city j. Specifically, a higher knowledge stock in city j would increase the overall level of spillovers from city j to city i. Second, another key component is the proximity between the two cities' technology fields. For example, if the structure of existing technology fields between two cities is closer, the technology spillovers from city j to city i will be greater (Bloom et al., 2013). Thus, the knowledge stock, together with the proximity of the two cities' technology fields, determines the level of technology spillovers between city i and city j.

There are two measures of technology spillovers: the Jaffe measure and the Mahalanobis measure. The key difference between the Jaffe measure and the Mahalanobis measure of technology spillovers lies in the assumption of whether there exists knowledge exchange across different technology fields. Specifically, the Jaffe measure assumes no knowledge exchange across different technology fields, focusing solely on spillovers within the same technology class. In contrast, the Mahalanobis measure accounts for both within-field and cross-field knowledge exchanges (Bloom et al., 2013).

We first construct the measure of technology spillovers proposed by Jaffe (1986). The Jaffe measure of technological spillover assumes that knowledge exchange only occurs within the same technology class, which rules out spillovers across different classes. Specifically, the Jaffe measure of technological spillover from city j to city i is specified as follows:

$$\text{SpillTech}_{ijt}^{\text{Jaffe}} = F_{it}\Phi F_{jt}' n_{it} n_{jt} \tag{3}$$

where $F_{it} = [\frac{n_{i1t}}{n_{it}}, \frac{n_{i2t}}{n_{it}}, \frac{n_{i\tau t}}{n_{it}}, ..., \frac{n_{iNt}}{n_{it}}]$ is an N-dimensional row vector representing the share of patents across technology classes τ of city i in year t. n_{it} represents the patent stock of city i in year t, which is calculated using a perpetual inventory method with a $\delta = 5\%$ depreciation rate.[4] Here, we use the number of patents to proxy for the level of knowledge stock. Since the Jaffe measure only takes into account spillovers within the same technology class, we define Φ as an identity matrix. A higher value of $\text{SpillTech}_{ijt}^{\text{Jaffe}}$ indicates larger technology spillovers from city j to city i.

We then calculate the magnitude of technology spillovers for city i in year t, using Equation (4). Intuitively, the summation of the technology spillovers from all other cities to city i captures the strength of technology spillover faced by city i. A higher value of $\text{SpillTech}_{i,t}^{\text{Jaffe}}$

[4]Specifically, knowledge stock is computed using the equation: $n_{j,t} = (1-\delta)n_{j,t-1} + Patent_{j,t}$, where $Patent_{j,t}$ is the newly filed patents in city j in year t and δ is the depreciation rate of 5%. 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 was established in 1985, and we can precisely estimate the patent stock for all cities from 1985.

indicates a greater amount of technology spillovers that city i faces in year t.

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

Second, we construct an alternative measure of technology spillovers, denoted as $\text{SpillTech}_{i,t}^{\text{Maha}}$, based on the Mahalanobis distance of technological proximity, as proposed by Bloom et al. (2013). This Mahalanobis measure captures technology spillovers both within and across technology classes.[5] Similar to the the Jaffe measure, a higher value of $\text{SpillTech}_{i,t}^{\text{Maha}}$ indicates larger technology spillovers faced by city i in year t. The detailed calculation of the Mahalanobis measure of technology spillovers, $\text{SpillTech}_{i,t}^{\text{Maha}}$, is presented in Appendix A.

City-Pair-Level Statistics. In addition, we construct the city-pair-level metrics of HSR connections and innovation activities to support our city-pair-level analyses. First, based on the evolution of the HSR network, we begin by identifying the earliest year that two cities, city i and city j, are connected by the HSR network, both directly and indirectly.[6] We then construct the city-pair-level dummy variable of HSR connections, $HSR_{ij,t}$, which takes a value of 1 if city i is connected to city j by any HSR line in year t and 0 otherwise.

Second, as patent backward citations measure technological knowledge flows from a citation-receiving patent to a citation-giving patent, inter-city patent citations capture knowledge diffusion from a patent-cited city to a patent-citing city (Jaffe et al., 2000). Therefore, we construct inter-city patent citations from city j to city i in year t to measure the flows of knowledge from cited-city j to citing-city i. Specifically, inter-city patent citations from city j to city i in year t is defined as the number of citations made by patents filed in city i in year t to prior patents filed in city j.

Third, given that inventors and assignees typically maintain close relationships and engage in extensive communications to share knowledge following a patent transfer, patent transactions emerge as a crucial conduit for knowledge diffusion (Liu et al., 2022). Therefore, we construct inter-city patent transactions from city j to city i, which is defined as the number of patent transactions sold from city j to city i. Specifically, since the authority requires that each patent transaction be registered with the CNIPA, we thus identify the buyer and seller for each patent transaction, as well as their corresponding locations, and construct the number of patents sold from city j to city i each year.

[5]Relative to Jaffe measure of technology spillover, Mahalanobis measure of technology spillover provides a more comprehensive assessment of technological relatedness

[6]For instance, they are directly 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 considered indirectly connected by HSR if the two HSR lines intersect in city k.

Fourth, as patent licenses grant the licensee permission to use the patent-protected knowledge held by the licensor, patent licenses serve as an appealing way through which knowledge diffusion takes place across cities (Nelson, 2009). Specifically, for each patent license agreement, we identify the patent licensor and licensee, along with their respective locations. We then define the inter-city patent licensing from city j to city i as the number of patents licensed from city j to city i.

Fifth, we define inter-city patent collaborations as the number of collaborative patents filed by applicants from city i and city j in year t. Identifying the addresses of collaborative patent applicants presents a significant challenge. While the patent data from CNIPA includes the names of all applicants in the patent filing document and the address of the first applicant, it does not provide the addresses of the remaining applicants. To address this limitation, we identify the addresses of the remaining applicants from the administrative data of firm registration. Therefore, we can identify all 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 classification, we compute the total number of research teams engaged in collaborative patents, the number of newly formed research teams, and the number of patents filed by those newly formed research teams.

Finally, we construct a measure of patent similarity between patents in city i and city j. We begin by calculating the pairwise cosine similarity between city i's patent vectors and city j's patent vectors in year t and for each technology class x, as specified in Equation (5):

$$\text{Similarity}_{ijx,t} = \frac{\sum_{p \in i,\ p' \in j} cosine(\vec{v}_{p,i,x,t}, \vec{v}_{p',j,x,t})}{N_{ixt} N_{jxt}} \tag{5}$$

where $cosine(\vec{v}_{p,i,x,t}, \vec{v}_{p',j,x,t})$ denotes the cosine similarity between city i's patent vectors $\vec{v}_{p,i,x,t}$ and city j's patent vectors $\vec{v}_{p',j,x,t}$ in year t for technology class x. Here, city i's patent vectors $\vec{v}_{p,i,x,t}$ and city j's patent vectors $\vec{v}_{p',j,x,t}$ are constructed based on patent texts using LLMs. N_{ixt} and N_{jxt} represent the number of patents in the technology class x in city i and city j, respectively. Next, we define inter-city patent similarity as the average pairwise cosine similarity for all technology classes.

3.3 Summary statistics

Table 1 presents the summary statistics of key variables for the city-level analysis. We find that the proportions of the top important patents in HSR-connected cities are larger than

those in unconnected cities. This pattern suggests that HSR-connected cities are often cities that undertake more breakthrough innovation, compared to unconnected cities. Moreover, the proportions of bottom important patents in HSR-connected cities are smaller than those in unconnected cities. In addition, the summary statistics also reveal that technology spillovers in HSR-connected cities are greater than those in unconnected cities, as supported by both the Jaffe measure and the Mahalanobis measure of technology spillovers.

Table 2 presents the summary statistics of key variables for the city-pair level analysis. As shown in Panel A, inter-city patent citations, transactions, and licensing activities are greater between city pairs connected by HSR, compared to those between unconnected city pairs. Furthermore, Panel B illustrates that the similarity of inter-city patents is greater between HSR-connected city pairs than between unconnected city pairs.

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:

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

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:

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

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 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.

5 The Impact of HSR Connections

5.1 Baseline Results

Table 3 reports the estimated effects of HSR connections on the composition of local innovation. We differentiate innovation 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 depends 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 3, 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 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).[7] 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.

[7]The sample mean of the proportion of the top 15% important patents is 0.115, and the sample mean of the proportion of the bottom 25% important patents is 0.368.

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 for its validity by testing the presence of pre-trends before the introduction of HSR connections. Figure 5 provides the event study graph for the effects of HSR connections. As shown in Figures 5(a)–5(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. This pattern suggests that the opening of HSR takes time to affect the breakthrough innovation. On the other hand, as Figures 5(d)–5(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 6 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-difference Regressions. 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 further 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.

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 regression. Specifically, as shown in Panel B of Table 4, HSR connections are positively correlated with the historical railway networks established in 1962, with the Kleibergen-Paap F-statistic exceeding the conventional threshold of 16. This result indicates that the historical railway network serves as a robust and relevant instrument for HSR connections, thereby mitigating concerns related to weak instrument bias. Furthermore, as the locations of the historical railway routes of 1962 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.

Panel A of Table 4 presents the long-difference regression results at the city level, using the instrumental variable approach. Specifically, in these long-difference specifications, we regress changes in the proportions of top and bottom important patents on the dummy variable of HSR connections indicating whether city i is connected to HSR networks during the period 2008–2017, using the historical railway network in 1962 as the instrument. Specifically, columns (1)–(3) in Panel A present the estimated effects of HSR connections on the changes in the proportion of top important patents. In all these columns, the estimated coefficients of HSR connections are negative, although the coefficients are imprecisely estimated in columns (1) and (2). The negative coefficients of HSR connections reconfirm 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. Although the estimated coefficients for HSR connections in columns (4)–(5) are imprecisely estimated, the estimated coefficients of HSR connections are close to the baseline estimates. These results suggest that HSR connections promote incremental innovation, which is consistent with our baseline findings. Overall, the results of the long-difference specifications confirm the asymmetric effects of HSR connections on incremental and breakthrough innovation in our baseline findings, suggesting

that the potential omitted variables do not bias our baseline findings.[8]

Placebo Tests We conduct placebo tests to examine whether our baseline findings have been driven by chance. Specifically, we randomly assign cities that launch HSR, set a pseudo year for the opening of HSR, and subsequently create a false HSR connections variable. Using these randomly assigned HSR connection data, we carry out the placebo estimation by replicating the baseline regressions 1000 times. Figure 7 reports the corresponding results for the proportion of top and bottom important patents, respectively. We find that the estimated coefficients of pseudo-HSR connections are centered around zero, indicating that there is no significant effect from the randomly assigned HSR connections. 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 5, 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 hindering breakthrough innovation, thereby leading to a structural shift 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 6. 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

[8]The results from long-difference specifications, however, should be interpreted with caution, as they are based on a small number of observations. In particular, with a small sample size, the statistical power of t-test decreases. That is, even if there is a true effect (i.e., a true relationship between HSR connections and the composition of local innovation), it is less likely to detect this true effect. As a result, the significance of regression coefficients in long-difference specifications can be attenuated, making it harder to achieve statistically significant results.

insignificant effects on the composition of incremental versus breakthrough innovation. These findings suggest that face-to-face interactions and subsequent tacit technology spillovers, facilitated by HSR connections, play a more significant role in shaping the composition of local innovation than online communication enabled by the advancement of ICT.

We also conduct a battery of robustness checks using alternative model specifications, which include province-specific linear time trends and additional fixed effects. Our main results remain robust to these alternative model specifications. Details of each robustness exercise are presented in Appendix C.

6 Mechanism: The Role of Technology Spillovers

To explore the mechanisms behind HSR's differential effects on incremental and breakthrough innovation, we begin by proposing a conceptual framework that illustrates how the composition of innovation responds to changes in the relative costs of innovation inputs resulting from the introduction of HSR. Guided by the conceptual framework, we then 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 explore how knowledge spillovers take place across cities, utilizing information on inter-city patent transactions, patent licensing, and patent collaborations.

6.1 A Conceptual Framework of Innovation

We present a conceptual model and its key features 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 patents with two inputs: (i) knowledge (both local and distant) and (ii) local capital (such as scientists, laboratories), combined via Cobb-Douglas aggregation. 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. Local capital inputs are costly to innovators, as they have to pay to employ scientists and establish laboratories. For simplicity, we assume a fixed capital endowment for each city.

The core of the model lies in the input composition difference between incremental and breakthrough innovation. Incremental innovation is knowledge-intensive and is heavily based on available knowledge stock. In contrast, breakthrough innovation is less knowledge-intensive but yields higher returns than incremental innovation (Manso, 2011). Following the introduction of more advanced transportation infrastructure (such as HSR connections), the geographic decay of knowledge decreases. Therefore, innovators have access to a larger pool of distant knowledge.

As a result, we expect to observe two outcomes following the introduction of HSR. First, as innovators gain access to a larger pool of distant knowledge, the expanded knowledge base, combined with the knowledge-intensive nature of incremental innovation, prompts firms to shift their efforts toward incremental projects (as formalized in Proposition 1 of Appendix B). In particular, the increased abundance of knowledge lowers the relative cost of incremental innovation compared to breakthrough innovation, thereby encouraging innovators to undertake more incremental projects. Second, facilitated knowledge flows produce a convergence effect. As transportation improves, the mix of innovative inputs (both knowledge and local 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 the role of knowledge spillovers, which are measured by cities' exposure to technology spillovers, and empirically examine their role in shifting the composition of local innovation. In Section 7, we discuss the direction of innovation between connected cities, which corresponds 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' exposure to technology spillovers, as measured by the Jaffe and Mahalanobis measures. Specifically, we separately regress Jaffe and Mahalanobis measures of technology spillovers on the dummy variable of HSR connections. Columns (1)–(2) of Table 7 report the regression results, in which the measures of technology spillovers are discounted at the 5% depreciation rate. 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. As additional robustness checks, we calculate the measures of technology spillovers using an alternative depreciation rate of 15% and present the estimation results in columns (3)–(4).[9]

[9]In the subsequent analysis, the Jaffe and Mahalanobis measures of technology spillovers are calculated with a depreciation rate of $\delta = 5\%$. Our findings remain robust when using an alternative depreciation rate

We find consistent results across different specifications. Together, these results demonstrate that HSR connections increase technology spillovers among HSR-connected cities. These findings also imply that HSR connections reshape the composition of local innovation by facilitating increased exposure to external technology 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 8, we separately regress the proportions of top and bottom important patents on the Jaffe measure of technology spillovers. The estimated coefficients of the HSR connections in columns (1)–(3) are all negative and significant, showing that technology spillovers discourage breakthrough innovation that explores unknown areas. In particular, a 10% increase in the Jaffe measure of technology spillovers is associated with a 1.7% decrease in the proportion of the top 15% important patents (relative to the sample mean). On the other hand, the coefficients of the HSR connections are positive in columns (4)–(6). These coefficients achieve statistical significance in column (4), although they are imprecisely estimated in columns (5) and (6). These results suggest that technology spillovers are more likely to stimulate incremental innovation that builds on existing knowledge. In sum, we find that technology spillovers promote more incremental innovation than breakthrough innovation.

In addition, in Panel B of Table 8, we separately regress the proportions of top and bottom important patents on the Mahalanobis measure of technology spillovers. The estimated results reveal a consistent pattern: technology 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 technology spillovers induced by HSR connections facilitate the exploitation of existing knowledge, which is beneficial to incremental innovation. However, these technology 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,

of $\delta = 15\%$.

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 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 connections 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. Specifically, we use patent citations between city i and city j as the dependent variable. $HSR_{ij,t-1}$ 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} is 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 9, 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 Does Technology Spill Over Across Cities?

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. Specifically, we perform this city-pair-level analysis using Equation (8), where the dependent variable is inter-city patent transactions between city i and city j. The regression results are presented in column (3) of Table 9. The estimated coefficient of the HSR connection variable is positive and statistically 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.[10] Thus, we further examine the ways through which

[10]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

knowledge spillovers occur by testing if HSR connections promote inter-city patent licensing. In particular, we explore this issue using Equation (8) and replace the dependent variable with inter-city patent licensing. Column (4) of Table 9 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 margins of inter-city patent collaborations, using the city-pair-level specification outlined in Equation (8).

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 10 present the estimation results. We find that HSR connections increase 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 increases by 5.27%. These results indicate that HSR connections facilitate patent collaborations at the intensive margin between pairs of cities connected by HSRs.

We then examine the effects of HSR connections on the extensive margin of inter-city patent collaborations. Columns (3) and (4) of Table 10 present the estimation results, where the dependent variables are the number of newly formed collaborative research teams and the number of patents filed by these newly formed research teams, respectively. Specifically, the result in column (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

include detailed laboratory reports, user manuals, on-site guidance, and technical training, which ensure that the licensee has easier access to both explicit and tacit knowledge.

margin.

Overall, the above results indicate that HSR connections lead to increased patent collaborations between HSR-connected cities, both at 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, and what the implications are for local innovation.

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 inter-city patent similarity between city i and city j as the dependent 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 connected city pairs.

Columns (1)-(2) of Table 11 present the estimation results, showing that the estimated coefficients of $\text{HSR}_{ij,t-1}$ 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).[11] 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-difference regressions at the city-pair level and report these regression results

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

in Table 12. Specifically, similar to the city-level IV analysis, we construct the corresponding city-pair level instrument, which is a dummy variable of whether cities i and j are connected by historical railway networks in 1962. Column (1) of Table 12 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-difference 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 greater inter-city technology spillovers between connected city pairs. To explore these issues, we regress the inter-city technology spillovers as measured by the Jaffe and Mahalanobis measures on the dummy variable of $\text{HSR}_{ij,t-1}$. As shown in columns (1) – (2) of Table 13, the estimated coefficients of $\text{HSR}_{ij,t-1}$ 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 the Jaffe and Mahalanobis measures. As shown in columns (3)-(4) of Table 13, greater inter-city technology spillovers are associated with increased patent similarity between connected cities. This finding indicates that inter-city technology spillovers induced by HSR connections shift the direction of innovation between connected cities toward greater similarity.

As inter-city patent similarity reflects the degree of convergence between cities beyond their existing technological capabilities, an increase in inter-city patent similarity implies that cities are increasingly engaged in similar innovative projects. Consequently, our results suggest that HSR connections steer the innovation trajectories of connected city pairs toward heightened similarity. Furthermore, we provide suggestive evidence that technology spillovers facilitated by HSR connections contribute to this convergence in the direction of innovation.

8 Conclusion

Previous studies have shown that advanced transportation infrastructure promotes innovation. However, most of these studies focus on overall innovation, typically measured by the quantity and quality of innovation outcomes. This study takes a different perspective by

exploring how transportation improvements reshape the composition, rather than merely the scale, of local innovation.

We empirically investigate the effects of advanced transportation on the composition of local innovation, using the expansion of China's HSR as a quasi-natural experiment. Using a large-scale dataset of Chinese patent filings from 2000 to 2017, we employ a large language model to refine the measures of important patents, providing more precise gauges of both 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 technology spillovers induced by HSR connections lead to asymmetric effects on incremental and breakthrough innovation. This occurs because increased technology spillovers reduce the relative cost in the exploitation of existing knowledge, compared to the exploration of unknown knowledge, thereby incentivizing connected cities to undertake more incremental innovation. Furthermore, we corroborate the presence of technology spillovers by showing that HSR connections enhance inter-city patent citations. When we investigate how technology spillovers across cities take place, we find evidence that technology spillovers are mainly driven by inter-city patent transactions, licensing, and collaborations.

In addition, we investigate how technology 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 technology spillovers between connected city pairs, which in turn significantly increases the between-city patent similarity. Together, these findings indicate that technology 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 technology spillovers, induced by HSR connectivity, influence the 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 example, governments ought to implement policies that counteract the downsides of high-speed rail connectivity, thereby creating an environment where breakthrough innovation can flourish.

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.

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.

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.

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

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.

Duranton, Gilles and Matthew A Turner, "Urban growth and transportation," *Review of Economic Studies*, 2012, *79* (4), 1407–1440.

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

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

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, "The NBER patent citation data file: Lessons, insights and methodological tools," *NBER Working Paper*, 2001.

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 Michael S Fogarty, "Knowledge spillovers and patent citations: Evidence from a survey of inventors," *American Economic Review*, 2000, *90* (2), 215–218.

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

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.

Lampe, Ryan, "Strategic citation," *Review of Economics and Statistics*, 2012, *94* (1), 320–333.

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

Lerner, Josh and Amit Seru, "The use and misuse of patent data: Issues for finance and beyond," *The Review of Financial Studies*, 2022, *35* (6), 2667–2704.

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.

_ , _ , Johan Sulaeman, Jubo Yan, and Jialiang Zhang, "Expanding footprints: The impact of passenger transportation on corporate locations," *Review of Finance*, 2023, *27* (3), 1119–1154.

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.

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.

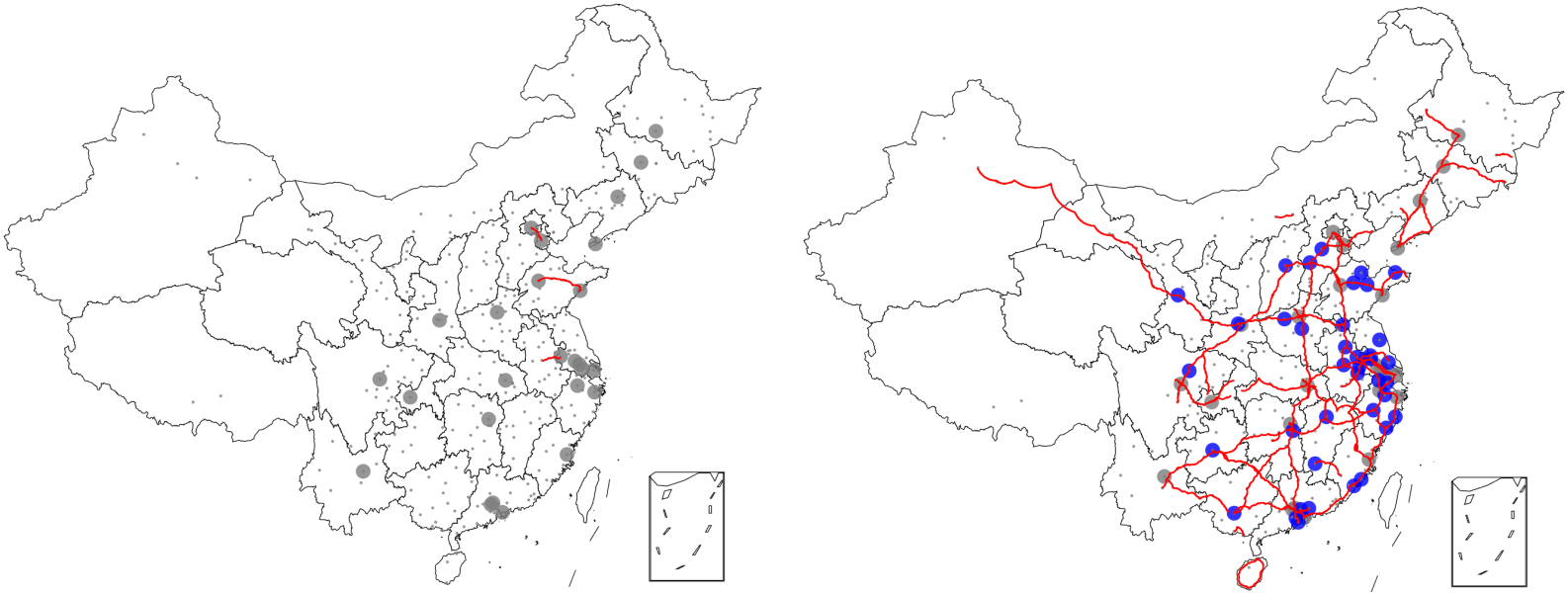
Thompson, Peter and Melanie Fox-Kean, "Patent citations and the geography of knowledge spillovers: A reassessment," *American Economic Review*, 2005, *95* (1), 450–460.

Figure 1: China's High-Speed Rail (HSR) Network (2008–2017)



Notes: This figure displays the geographic distribution of HSR routes in China from 2008 to 2017. The orange lines correspond to the HSR routes that were established from 2008 to 2012. The navy lines represent the HSR routes that were constructed between 2013 and 2017.

Figure 2: HSR Routes and Cities with Above-average Patent Filings in 2008 and 2017

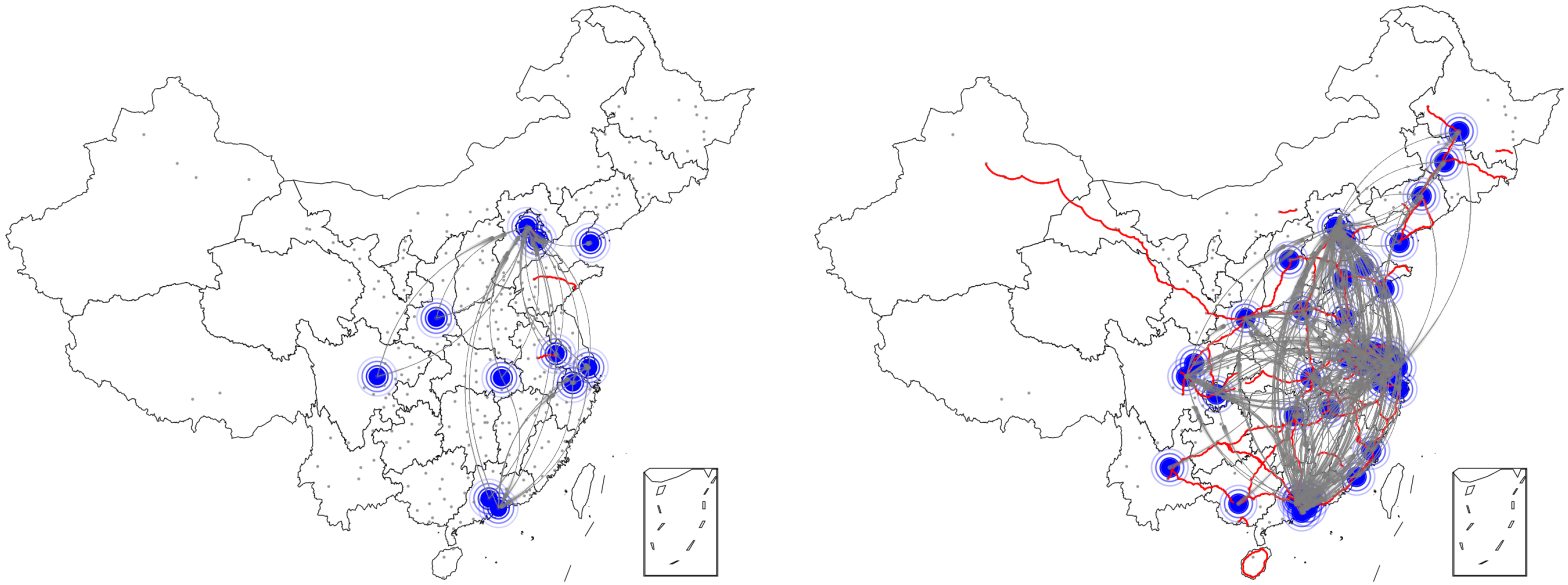


(a) Cities with above-average patent filings in 2008

(b) Cities with above-average patent filings in 2017

Notes: Figure 2(a) displays the geographic distribution of 36 cities with above-average (granted) patent filings in 2008, where the gray dots represent the locations of the 36 cities and the red lines denote the HSR routes. Figure 2(b) displays the geographic distribution of 56 cities with above-average (granted) patent filings in 2017. The blue dots represent the locations of the 20 newly emerged highly innovative cities whose patent filings exceeded the average in 2017, while the gray dots denote the 36 cities whose patent filings exceeded the average in 2008. The red lines denote the HSR routes.

Figure 3: HSR Routes and Cities with Above-average Inter-city Patent Citations in 2008 and 2017

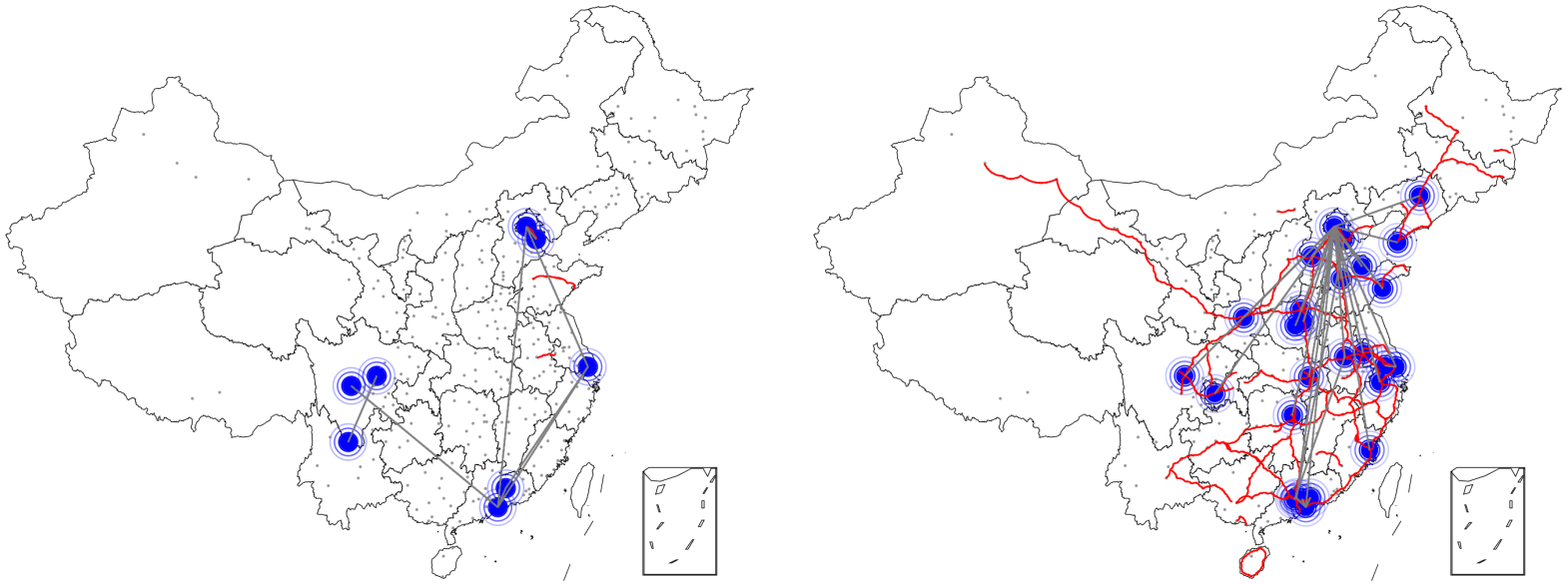


(a) Cities with above-average inter-city citations in 2008

(b) Cities with above-average inter-city citations in 2017

Notes: Figure 3(a) displays the geographic distribution of cities with above-average inter-city citations in 2008. In this figure, the blue dots indicate the locations of these cities. The gray lines represent the collaborative relationships between these cities, while the red lines denote the HSR routes. Figure 3(b) displays the geographic distribution of cities with above-average inter-city citations in 2017. In this figure, the blue dots also represent the locations of the corresponding cities, the gray lines depict collaborative relationships between connected city pairs, and the red lines indicate the HSR routes.

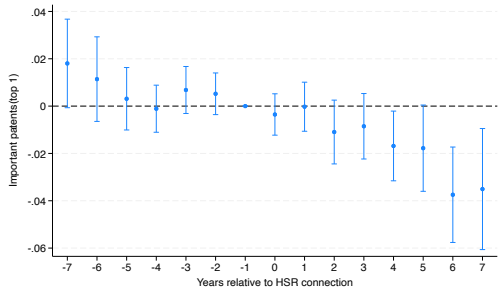
Figure 4: HSR Routes and Cities with Above-average Inter-city Collaborative Patents in 2008 and 2017



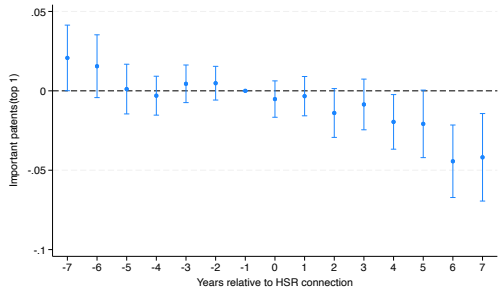
(a) Cities with above-average inter-city collaborative patents in 2008 (b) Cities with above-average inter-city collaborative patents in 2017

Notes: Figure 4(a) displays the geographic distribution of cities with above-average inter-city collaborative patents in 2008. In this figure, the blue dots indicate the locations of these cities. The gray lines represent the collaborative relationships between these cities, while the red lines denote the HSR routes. Figure 4(b) displays the geographic distribution of cities with above-average inter-city patent collaborations in 2017. In this figure, the blue dots also represent the locations of the corresponding cities, the gray lines depict collaborative relationships between connected city pairs, and the red lines indicate the HSR routes.

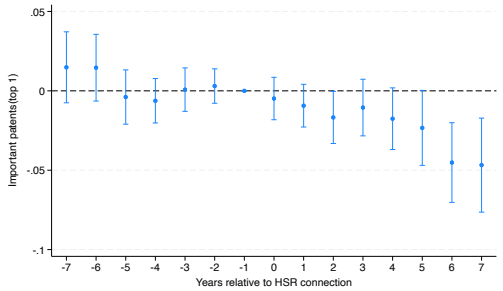
Figure 5: Event Study Graphics: Impact of HSR Connections on Incremental vs. Breakthrough Innovation



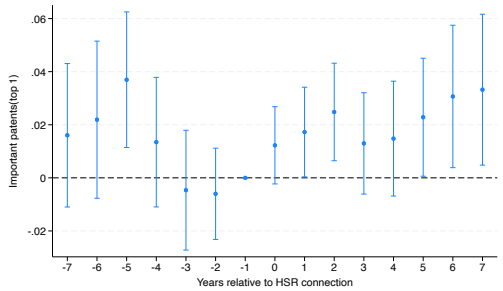
(a) Top Important Patents(15%)



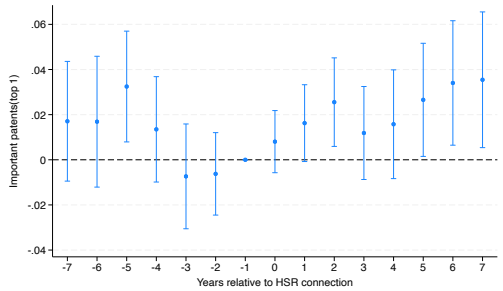
(b) Top Important Patents(20%)



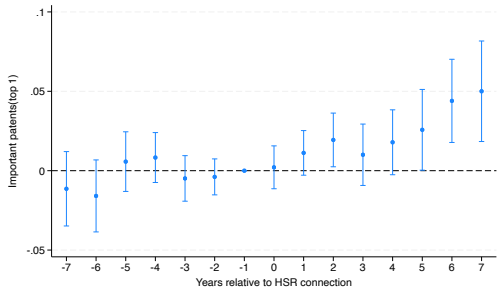
(c) Top Important Patents(25%)



(d) Bottom Important Patents(25%)



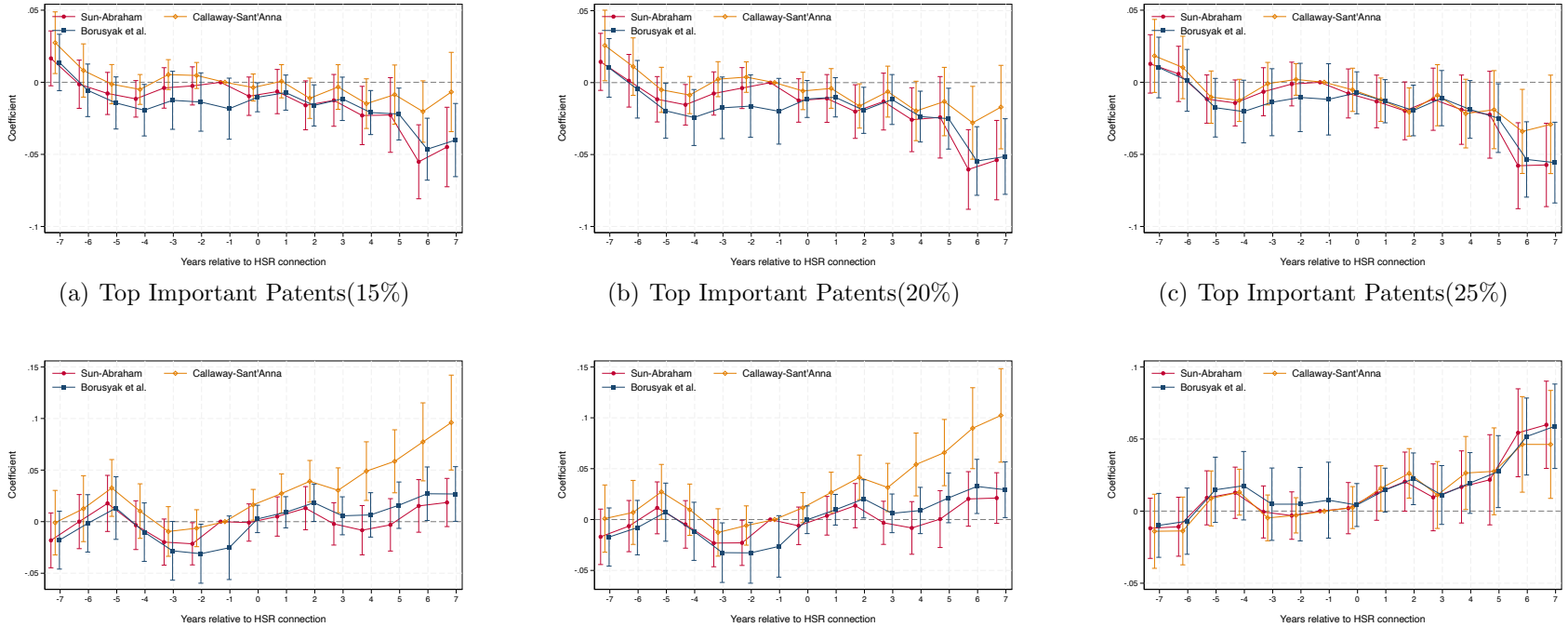
(e) Bottom Important Patents(30%)



(f) Bottom Important Patents(70%)

Notes: This figure presents the estimated effects of HSR connections on incremental and breakthrough innovation, using event study specifications. The dependent variables in Figures (a)-(c) are the proportions of top 15%, top 20%, and top 25% important patents, respectively, which represent breakthrough innovation. The dependent variables in Figures (d)-(f) are the proportions of bottom 25%, bottom 30%, and bottom 70% important patents, respectively, which represent incremental innovation. The blue bars indicate 95% confidence intervals around the estimated dynamic coefficients (blue dots).

Figure 6: Event Study Graphics: using Robust DiD Estimators



(a) Top Important Patents(15%)

(b) Top Important Patents(20%)

(c) Top Important Patents(25%)

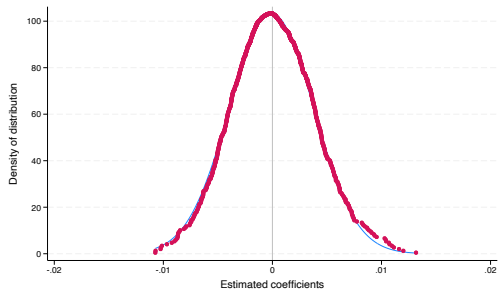
(d) Bottom Important Patents(25%)

(e) Bottom Important Patents(30%)

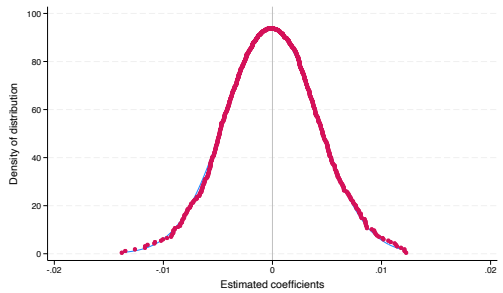
(f) Bottom Important Patents(70%)

Notes: This figure presents the estimated effects of HSR connections on incremental and breakthrough innovation, using a set of robust DiD estimators introduced by Sun and Abraham (2021), Callaway and Sant'Anna (2021), and Borusyak et al. (2024). The dependent variables in Figures (a)-(c) are the proportions of top 15%, top 20%, and top 25% important patents, respectively, which represent breakthrough innovation. The dependent variables in Figures (d)-(f) are the proportions of bottom 25%, bottom 30%, and bottom 70% important patents, respectively, which represent incremental innovation. In the figure, the dotted red lines correspond to the estimation results of the robust DiD estimator developed by Sun and Abraham (2021). The diamond-shaped yellow lines reflect the results from the estimator by Callaway and Sant'Anna (2021). The squared blue lines depict the estimation results from the estimator introduced by Borusyak et al. (2024). The bars indicate 95% confidence intervals around the estimated dynamic coefficients.

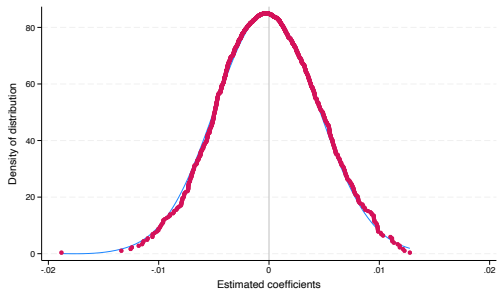
Figure 7: Distribution of Estimates in the Placebo Test (City Level)



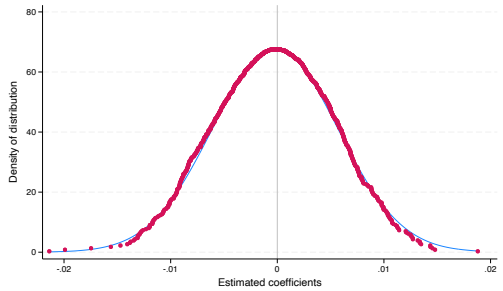
(a) Top Important Patents(15%)



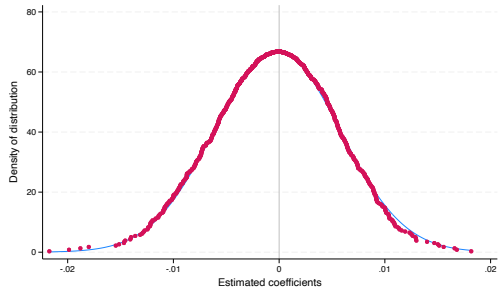
(b) Top Important Patents(20%)



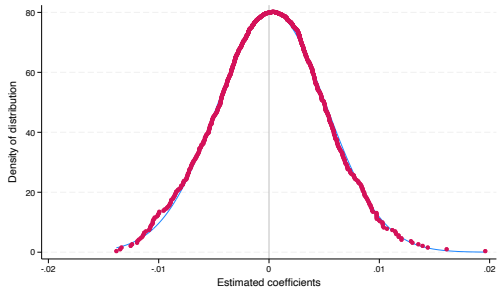
(c) Top Important Patents(25%)



(d) Bottom Important Patents(25%)



(e) Bottom Important Patents(30%)



(f) Bottom Important Patents(70%)

Notes: This figure shows the distribution of the Placebo estimations. The dependent variables in Figures (a)-(c) are the proportions of top 15%, top 20%, and top 25% important patents, respectively, which represent breakthrough innovation. The dependent variables in Figures (d)-(f) are the proportions of bottom 25%, bottom 30%, and bottom 70% important patents, respectively, which represent incremental innovation. The explanatory variable is the pseudo-HSR connections. We replicate the baseline regressions 1000 times to obtain the distribution of estimated coefficients of pseudo-HSR connections.

Table 1: Summary Statistics: City-level Variables

	Full sample	HSR-connected cities	Unconnected cities
	(1)	(2)	(3)
Important patents(top 15%)	0.11	0.13	0.11
	(0.12)	(0.07)	(0.13)
Important patents(top 20%)	0.15	0.18	0.14
	(0.14)	(0.09)	(0.15)
Important patents(top 25%)	0.18	0.23	0.17
	(0.15)	(0.10)	(0.16)
Important patents(bottom 25%)	0.37	0.26	0.41
	(0.22)	(0.08)	(0.24)
Important patents(bottom 30%)	0.42	0.31	0.45
	(0.22)	(0.09)	(0.24)
Important patents(bottom 70%)	0.78	0.73	0.80
	(0.16)	(0.10)	(0.17)
Invention patent counts	600.09	2130.92	181.23
	(2792.57)	(5612.95)	(708.34)
Patent citation counts	1321.56	4346.63	493.83
	(6473.33)	(12789.38)	(2336.60)
Log(GDP per capita)	10.00	10.83	9.77
	(0.93)	(0.69)	(0.85)
Log(wage)	10.17	10.84	9.98
	(0.67)	(0.29)	(0.62)
% GDP of secondary industry	49.85	48.07	50.34
	(12.14)	(9.98)	(12.63)
% GDP of tertiary industry	42.69	46.87	41.55
	(10.86)	(10.64)	(10.64)
Road ridership	8.45	8.81	8.35
	(0.95)	(1.01)	(0.91)
Air ridership	1.66	2.58	1.41
	(2.33)	(2.99)	(2.04)
Broadband Subscriber	495.77	1235.99	295.51
	(776.16)	(1241.90)	(400.13)
Technology spillovers (Jaffe)	15.27	18.23	14.46
	(2.56)	(1.87)	(2.08)
Technology spillovers (Mahalanobis)	15.32	18.26	14.52
	(2.54)	(1.85)	(2.07)
Observations	4664	1002	3662

Standard deviations in parentheses.

Notes: This table presents the summary statistics for key variables constructed based on city-level panel data. Column (1) reports the mean and standard deviation of the key variables for the full sample, while columns (2) and (3) present the corresponding statistics for HSR-connected and unconnected cities, respectively.

Table 2: Summary Statistics: City-pair Level Variables

	Full sample	HSR-connected city pairs	Unconnected city pairs
	(1)	(2)	(3)
Panel A			
Inter-city patent citations (university patents)	0.25	1.06	0.18
	(0.70)	(1.41)	(0.53)
Inter-city citations of university patents	0.10	0.52	0.06
	(0.47)	(1.09)	(0.33)
Inter-city patent transactions	0.02	0.16	0.01
	(0.20)	(0.52)	(0.12)
Inter-city patent licensing	0.01	0.03	0.00
	(0.08)	(0.19)	(0.07)
Inter-city similarity	0.07	0.21	0.06
	(0.09)	(0.12)	(0.07)
Inter-city technology spillovers (Jaffe)	8.19	11.94	7.84
	(2.81)	(2.30)	(2.59)
Inter-city technology spillovers (Mahalanobis)	8.22	11.99	7.86
	(2.83)	(2.30)	(2.61)
Sum of two cities' GDP per capita	10.07	10.97	9.98
	(0.84)	(0.56)	(0.81)
Sum of two cities' average wage	10.19	10.93	10.12
	(0.63)	(0.21)	(0.61)
Sum of two cities' % GDP of secondary industry	99.52	94.52	99.98
	(17.60)	(14.17)	(17.81)
Sum of two cities' % GDP of tertiary industry	92.44	94.97	92.21
	(16.03)	(13.94)	(16.20)
Observations	1317968	112544	1205424

Standard deviations in parentheses.

Notes: This table presents the summary statistics for key variables constructed based on city-pair level panel data. Column (1) reports the mean and standard deviation of the key variables for the full sample, while columns (2) and (3) present the corresponding statistics for HSR-connected and unconnected city pairs, respectively. Inter-city patent citations, transactions, and licensing are directional. For example, inter-city patent citations from Beijing to Shanghai and inter-city patent citations from Shanghai to Beijing indicate knowledge flows in opposite directions. Inter-city patent similarity and the Jaffe measure of technology spillovers are undirectional, while the Mahalanobis measure of technology spillovers is directional.

Table 3: 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 connections	-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.

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

Notes: This table reports the impact of HSR connections on incremental and breakthrough innovation, respectively. Specifically, columns (1) - (3) present the impact of HSR connections on breakthrough innovation, where the dependent variables are the proportions of top 15%, top 20%, and top 25% important patents, respectively. Columns (4)-(6) present the impact of HSR connections on incremental innovation, where the dependent variables are the proportions of bottom 25%, bottom 30%, and bottom 70% important patents, respectively. Time-varying city-level controls include GDP per capita, average wage, percentage of GDP from the secondary industry, and percentage of GDP from the tertiary industry. All columns control for city fixed effects and year fixed effects. Standard errors are clustered at the city level.

Table 4: Impact of HSR Connections 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%
Panel A: IV = historical railroad						
HSR connections (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)
Panel B: first-stage regression; dep. var. = HSR connections						
Historical Railroad						0.356***
						(5.71)
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.

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

Notes: This table reports the estimated effects of HSR connections on incremental and breakthrough innovation using long difference IV regressions. Panel A presents the second-stage regressions using the historical railroad in 1962 as the instrumental variables. In columns (1)-(3), the dependent variables are the first-differenced proportions of top 15%, top 20%, and top 25% important patents between 2007 and 2017. In columns (4)-(6), the dependent variables are the first-differenced proportions of bottom 25%, bottom 30%, and bottom 70% important patents between 2007 and 2017. Panel B reports the first-stage regression, where the dependent variable is the dummy variable of HSR connections, indicating whether city i is connected during the period of 2008–2017. The explanatory variable is the historical railroad network in 1962. City-level controls include (first-differenced) GDP per capita, average wage, percentage of GDP from the secondary industry, and percentage of GDP from the tertiary industry. All columns control for province fixed effects. Standard errors are clustered at the city level.

Table 5: 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 connections	-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.

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

Notes: This table reports the estimates of the effects of HSR connections on incremental and breakthrough innovation when controlling for the impact of road and air ridership. Specifically, in all columns, we control for the number of road passengers and air passengers. City-level controls include GDP per capita, average wage, percentage of GDP from the secondary industry, and percentage of GDP from the tertiary industry. All columns control for city fixed effects and year fixed effects. Standard errors are clustered at the city level.

Table 6: 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 connections	-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.

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

Notes: This table reports the estimates of the effects of HSR connections on incremental and breakthrough innovation when controlling for the development of ICT. Specifically, in all columns, we control for the logarithm of one plus the number of broadband subscribers, which is used to measure the ICT development. City-level controls include GDP per capita, average wage, percentage of GDP from the secondary industry, and percentage of GDP from the tertiary industry. All columns control for city fixed effects and year fixed effects. Standard errors are clustered at the city level.

Table 7: The Impact of HSR connections on Technology Spillovers

	Technology spillovers ($\delta = 5\%$)		Technology spillovers ($\delta = 15\%$)	
	(1)	(2)	(3)	(4)
	Jaffe measure	Mahalanobis measure	Jaffe measure	Mahalanobis measure
L.HSR connections	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.

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

Notes: This table reports the estimated effects of HSR connections on technology spillovers. In columns (1) and (2), the dependent variables are the Jaffe measure of technology spillovers and the Mahalanobis measure of technology spillovers, respectively, where the depreciation rates are 5%. In columns (3) and (4), the dependent variables are the Jaffe measure of technology spillovers and the Mahalanobis measure of technology spillovers, respectively, where the depreciation rates are 15%. City-level controls include GDP per capita, average wage, percentage of GDP from the secondary industry, and percentage of GDP from the tertiary industry. All columns control for city fixed effects and year fixed effects. Standard errors are clustered at the city level.

Table 8: The Role of Technology Spillovers in Shaping 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%
Panel A: Jaffe measure						
L.SpillTech (Jaffe)	-0.0195***	-0.0214***	-0.0220***	0.00326	0.00604	0.0214***
	(-3.47)	(-3.48)	(-3.34)	(0.52)	(0.96)	(2.95)
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	0.364	0.409	0.437	0.602	0.594	0.467
Adj.R^2	0.319	0.367	0.398	0.574	0.565	0.430
Panel B: Mahalanobis measure						
L.SpillTech (Mahalanobis)	-0.0163***	-0.0172***	-0.0200***	0.00380	0.00661	0.0193**
	(-3.11)	(-3.04)	(-2.97)	(0.60)	(1.04)	(2.59)
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	0.334	0.363	0.436	0.602	0.594	0.467
Adj.R^2	0.287	0.318	0.397	0.574	0.565	0.429

t statistics in parentheses.

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

Notes: This table reports the estimated effects of technology spillovers on incremental and breakthrough innovation, respectively. In Panel A, the explanatory variables of interest are the lagged Jaffe measure of technology spillovers. In Panel B, the explanatory variables of interest are the lagged Mahalanobis measure of technology spillovers. Here, the depreciation rates are 5%. In both Panels A and B, the dependent variables in columns (1)-(3) are the proportions of top 15%, top 20%, and top 25% important patents, whereas the dependent variables in columns (4)-(6) are the proportions of bottom 25%, top 30%, and bottom 70% important patents. City-level controls include GDP per capita, average wage, percentage of GDP from the secondary industry, and percentage of GDP from the tertiary industry. All columns control for city fixed effects and year fixed effects. Standard errors are clustered at the city level.

Table 9: Inter-city Patent Citations, Transactions, and Licensing (city-pair level)

	Inter-city citations		Patent transactions	Patent licensing
	(1)	(2)	(3)	(4)
	Inter-city citations	Inter-city citations of university patents	Patent transactions	Patent licensing
L.HSR connections	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.

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

Notes: This table reports the estimated effects of HSR connections on inter-city patent citations, transactions, and licensing, using city-pair level analysis. The explanatory variable of interest is $HSR_{ij,t-1}$, which indicates whether city i and city j are connected by HSR in year $t-1$. The dependent variables in columns (1) and (2) are (logarithm of one plus) the number of inter-city patent citations and inter-city citations of university patents, respectively. The dependent variable in column (3) is (logarithm of one plus) the number of inter-city patent transactions. The dependent variable in column (4) is (logarithm of one plus) the number of patent licensing. City-pair controls include the sum of two cities' GDP per capita, the sum of two cities' average wage, the sum of two cities' percentage of GDP from secondary industry, and the sum of two cities' percentage of GDP from the tertiary industry. All columns control for city-pair fixed effects, citing-city by year fixed effects, and cited-city by year fixed effects. Standard errors are clustered at the city-pair level.

Table 10: 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 connections	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.

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

Notes: This table reports the estimated effects of HSR connections on inter-city patent collaborations, using the city-pair level analysis. The explanatory variable of interest is $HSR_{ij,t-1}$, which indicates whether city i and city j are connected by HSR in year $t-1$. Columns (1) and (2) examine the impact of HSR connections on collaborations at the intensive margin. The corresponding dependent variables are (logarithm of one plus) the number of inter-city collaborative patents and the number of inter-city collaborative research teams, respectively. Columns (3) and (4) examine the impact of HSR connections on collaborations at the extensive margin. The corresponding dependent variables are (logarithm of one plus) the number of new collaborative research teams and the number of patents filed by these new collaborative research teams, respectively. City-pair controls include the sum of two cities' GDP per capita, the sum of two cities' average wage, the sum of two cities' percentage of GDP from secondary industry, and the sum of two cities' percentage of GDP from the tertiary industry. All columns control for city-pair fixed effects, citing-city by year fixed effects, and cited-city by year fixed effects. Standard errors are clustered at the city-pair level.

Table 11: Effects of HSR Connections on Direction of Innovation across Cities

	Direction of Technological Innovaton	
	(1)	(2)
	Similarity	Similarity
L.HSR connections	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.

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

Notes: This table reports the estimated effects of HSR connections on inter-city patent similarity, using city-pair level specifications. The dependent variables in all columns are inter-city patent similarity between city i and city j. The explanatory variable of interest is $HSR_{ij,t-1}$, which indicates whether city i and city j are connected by HSR in year $t-1$. Column (1) includes no city-pair controls, whereas column (2) includes city-pair controls. City-pair controls include the sum of two cities' GDP per capita, the sum of two cities' average wage, the sum of two cities' percentage of GDP from secondary industry, and the sum of two cities' percentage of GDP from the tertiary industry. All columns control for city-pair fixed effects, citing-city by year fixed effects, and cited-city by year fixed effects. Standard errors are clustered at city-pair level.

Table 12: Effects of HSR connections on Inter-city Similarity: Long Difference IV Regressions

	Long-differenced IV	First-stage
	(1)	(2)
	Inter-city similarity	HSR connection
HSR connections	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.

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

Notes: This table reports the estimated effects of HSR connections on inter-city patent similarity, using city-pair level long difference IV regressions. The IV is the historical railroad network, which indicates whether city i and city j were connected by the historical railroad in 1962. Column (1) presents the second-stage estimation results using the 2SLS method, whereas column (2) presents the first-stage estimation results. The Kleibergen-Paap F-statistics is 305.780, which is larger than the threshold value of 10, indicating the historical railroad network is not a weak instrument. City-pair controls include the sum of two cities' GDP per capita, the sum of two cities' average wage, the sum of two cities' percentage of GDP from secondary industry, and the sum of two cities' percentage of GDP from the tertiary industry. All columns control for province-pair fixed effects. Standard errors are clustered at province-pair level.

Table 13: 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***	0.0135***		
	(6.59)	(10.52)		
L.Intercity technology spillovers(Jaffe measure)			0.00857***	
			(25.58)	
L.Intercity technology spillovers(Mahalanobis measure)				0.0115***
				(26.90)
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	1317968	658832	1317968
R^2	0.998	0.999	0.968	0.966
Adj.R^2	0.997	0.999	0.966	0.964

t statistics in parentheses.

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

Notes: Columns (1) and (2) in this table report estimated effects of HSR connections on inter-city technology spillovers, using city-pair level analysis. Specifically, the dependent variables in columns (1) and (2) are the Jaffe measure of technology spillovers and Mahalanobis measure of technology spillovers, respectively. The explanatory variables in columns (1) and (2) are $HSR_{ij,t-1}$, which indicates whether city i and city j are connected by HSR in year $t-1$. Columns(3) and (4) report the impact of inter-city technology spillovers on inter-city patent similarity. Specifically, the dependent variables in columns (3) and (4) are inter-city patent similarity. The explanatory variables in columns (3) and (4) are lagged Jaffe measure of inter-city technology spillovers and lagged Mahalanobis measure of inter-city technology spillovers, respectively. City-pair controls include the sum of two cities' GDP per capita, the sum of two cities' average wage, the sum of two cities' percentage of GDP from secondary industry, and the sum of two cities' percentage of GDP from the tertiary industry. All columns control for city-pair fixed effects, citing-city by year fixed effects, and cited-city by year fixed effects. Standard errors are clustered at the city-pair level.

Online Appendix

A Mahalanobis Measure of Technology Spillovers OA-1

B Conceptual Framework OA-3

C Additional Robustness Checks OA-6

A Mahalanobis Measure of Technology Spillovers

Unlike the Jaffe measure of technology spillovers, which only considers knowledge exchange within the same technology field, the Mahalanobis measure allows for both within- and cross-knowledge exchange (Bloom et al., 2013). For example, some technology classes, such as Chemistry, Physics, and Electricity, are important sources of knowledge in other fields. Thus, the Mahalanobis measure evaluates the knowledge spillovers in a more comprehensive way. Specifically, the Mahalanobis measure of inter-city technology spillovers from city j and city i is specified as follows:

$$\text{SpillTech}_{ijt}^{\text{Maha}} = F_{it}\Omega F_{jt}' n_{it} n_{jt} \tag{9}$$

where F_{it} is a N-dimensional row vector representing the share of patents across technology classes of city i in year t, and n_{it} is the patent stock of city i in year t measured by the perpetual inventory method with a $\delta = 5\%$ depreciation rate. A higher value of $\text{SpillTech}_{ijt}^{\text{Maha}}$ indicates larger technology spillovers from city j to city i.

Since the Mahalanobis measure takes account of the knowledge exchange within and across technology classes, we define Ω as the probability matrix of knowledge transfer across technology classes. To obtain a proper measure of the probability of knowledge transfer across technology classes, we utilize backward citation data from US patents. We collect the granted patents that were filed between 2000 and 2008 in the US Patent Office and their backward citations. The citing patents and cited patents are aggregated by their technology class, and we then calculate the share of backward citations over technology classes.

The matrix Ω is then defined in the following way. It is an N by N matrix where N is the total number of technology classes, and the rows represent the citing technology class, while the columns represent the cited technology classes. In the cell $\Omega_{m,n}$ of the matrix, we calculate the share of backward citations from technology m made to technology class n. As a result, Equation (9) represents the technology spillover from city j to city i.

We then calculate the magnitude of technology spillovers for city i in year t, which is based on the Mahalanobis distance. Specifically, 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. The specification of the Mahalanobis measure of technology spillovers that city i faces is as follows:

$$\text{SpillTech}_{i,t}^{\text{Maha}} = \sum_{i \neq j} \text{SpillTech}_{ijt}^{\text{Maha}} \tag{10}$$

Intuitively, a higher value of $\text{SpillTech}_{i,t}^{\text{Maha}}$ indicates a greater amount of technology spillovers faced by city i in year t.

B Conceptual Framework

We construct a conceptual framework to illustrate the different effects of infrastructure on two types of innovation, 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 innovation. 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 innovation. 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 innovation 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 innovation. We consider two measurements to reflect major aspects of local innovation in our model. First, we use the ratio of incremental to breakthrough innovation to represent the composition of local innovation:

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

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 innovation. 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 innovation

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 innovation 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 innovation in two cities.

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

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.

C Additional Robustness Checks

There is concern that some province-level characteristics, such as geographic features, land use policies, and industrial policies, may affect the local innovation activities. To capture any linear changes to province-level characteristics, we include province-specific linear time trends. As reported in Panel A of Table OA1, after controlling for province-specific linear time trends, our main results that HSR connections discourage breakthrough innovation while promoting incremental innovation remain robust. Furthermore, to control for any province-level policy changes that affect all cities within the same province equally, we include the province-by-year fixed effects. As shown in Panel B of Table OA1, the results are consistent with our main findings, although the coefficients in columns (4) and (5) are imprecisely estimated.

Table OA1: Robustness: inclusion of additional linear time trends and fixed effects

	Top Important Patents			Bottom Important Patents		
	(1)	(2)	(3)	(4)	(5)	(6)
	Top 15%	Top 20%	Top 25%	Bottom 25%	Bottom 30%	Bottom 70%
Panel A: Controlling for province-specific linear time trends						
L.HSR connections	-0.0105**	-0.0113*	-0.0124*	0.0133*	0.0155**	0.0148**
	(-2.03)	(-1.86)	(-1.87)	(1.90)	(2.11)	(2.11)
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
Province-specific linear time trends	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Controlling for province-by-year fixed effects						
L.HSR connections	-0.0125**	-0.0120*	-0.0119*	0.00843	0.0101	0.0133*
	(-2.26)	(-1.84)	(-1.66)	(1.12)	(1.30)	(1.75)
City-level Controls	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Province-by-year FE	Yes	Yes	Yes	Yes	Yes	Yes

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

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

Notes: This table reports the estimated effects of HSR connections on the composition of local innovation when controlling for additional linear time trends and fixed effects. Specifically, regressions in Panel A include the province-specific linear time trends. Regressions in Panel B include the province-by-year fixed effects.