High-Speed Rail and Scientific Collaboration: Evidence from China

Daiwei Chen & Pierre-Alexandre Balland

Papers in Evolutionary Economic Geography

26.05

Utrecht University Human Geography and Planning

High-Speed Rail and Scientific Collaboration: Evidence from China

Daiwei Chen[1] Pierre-Alexandre Balland[2,3]

April 27 2026

Abstract

China's high-speed rail (HSR) network, initiated in 2008, now covers nearly all regions of the country. This paper analyzes the effect of HSR connection on inter-city scientific collaboration and examines whether this effect varies systematically with the complexity of scientific fields. Combining the universe of HSR openings between 2008 and 2020 with OpenAlex publication records, we construct a panel spanning 33,793 Chinese city pairs. Using a staggered difference-in-differences estimator, we find that HSR increases co-publications among city-pairs with existing collaborative ties by 35.2 percent at the city-pair level. Disaggregating across twenty scientific fields, we show that this effect is quite heterogeneous. Field-level treatment effects range from 19.8 to 45.1 percent, and their magnitude is positively and significantly correlated with average team size - a proxy of the fields' complexity. These results are consistent with the view that face-to-face interaction is still important for knowledge production requiring deep divisions of cognitive labour, and they carry direct implications for the design of transportation and innovation policy.

Keywords: High-Speed Rail (HSR), Scientific Collaboration, Knowledge Complexity, Face-to-Face Interaction.

JEL Codes: O33, O38, R11, R58.

[1]University of Virginia, Department of Economics, Charlottesville, VA, USA
[2]Centre for European Policy Studies (CEPS), Brussels, Belgium.
[3]Harvard University, Growth Lab, Cambridge, MA, USA.

1 Introduction

Transportation infrastructure has long been viewed as an instrument of economic development, and its role in shaping the geography of innovation has attracted growing attention in recent decades (Aschauer 1989; Munnell 1990; Agrawal et al. 2017). Nowhere is this visible on a larger scale than in China, where the opening of the Beijing–Tianjin Intercity Railway in 2008 marked the start of what has since become the world's largest high-speed rail (HSR) network. By 2020, the system connected 284[1] cities and had fundamentally reshaped the cost of inter-city mobility, altering the opportunity set available to Chinese researchers, firms, and institutions.

A growing empirical literature documents the innovation consequences of this infrastructure push. Using patent-based measures of innovative output, existing studies broadly find positive effects of HSR connection on both the quantity and the quality of local innovation (Zheng et al. 2022; Wang and Cai 2020; Hanley et al. 2021). Kang et al. (2023) extend this result to city pairs that are indirectly connected through the HSR network, while Cui et al. (2020) document gains in university–industry co-patenting. Evidence based on research publications is scarcer but similarly supportive: Dong et al. (2020) find that HSR connection raises scientific co-publication and induces the migration of productive scientists toward connected cities. Taken together, this body of work establishes a robust positive relationship between HSR access and measured innovative activity.

A parallel literature in economic geography has shifted attention from the quantity of innovation to the nature of the knowledge being produced. Work in economic complexity argues that not all knowledge is equally productive: some forms of knowledge are systematically more important for long-run growth than others (Balland and Rigby 2017). Balland et al. (2020) show that complex economic activities concentrate in large urban centres, and Balland et al. (2019) argue that regional diversification into related but more complex tech-

[1] In this paper we focus on 270 cities, which specifically constitute Chinese prefecture-level cities in the HSR system.

nological domains is central to effective smart specialisation. Li and Rigby (2022) extend this framework to Chinese patent data, documenting a significant association between diversification into related and complex technologies and city-level GDP growth. Applied to the HSR setting, Chen and Guo (2023) find that HSR promotes technological specialisation, while Dong et al. (2025) - to our knowledge the only study applying standard complexity indices in this context - show that HSR enables cities to diversify into less related and more complex technological domains.

These two literatures speak to one another in a way that has not been formally tested. If HSR facilitates innovation by lowering the cost of face-to-face interaction, and if the production of complex knowledge depends disproportionately on coordination among increasingly specialised individuals (Agrawal et al. 2017; Balland et al. 2020), then the collaborative gains from HSR connection should themselves be heterogeneous: larger where the underlying knowledge is more complex, smaller where it is not. The findings of Dong et al. (2025) on the output side naturally invite an examination of the upstream collaborative process through which complex knowledge is produced. To our knowledge, this mechanism has not been directly tested.

This paper fills that gap. We ask two related questions. First, does HSR connection between two Chinese cities raise the volume of scientific collaboration between researchers based in those cities? Second, and more importantly, does the HSR collaboration relationship vary systematically with the complexity of the underlying scientific field? To answer these questions we assemble a novel panel combining HSR openings between 2008 and 2020 with publication records from OpenAlex covering 2000 to 2025, from which we derive annual connection status for roughly 970,000 city pairs and annual co-publication counts at the city-pair–field level (Table 1). We proxy field complexity by the average number of collaborators per publication within a field, a measure motivated by the team-science literature (Jones 2009; Wuchty, Jones, and Uzzi 2007) and discussed in detail in Section 3.[2] We estimate the

[2]We use the average number of co-authors per publication to proxy the relative complexity of each research domain rather than complexity indices drawn directly from the economic complexity literature. We

causal effect of HSR connection on co-publication using the staggered difference-in-differences estimator of Callaway and Sant'Anna (2021), and relate the estimated field-level effects to our complexity proxy through ordinary least squares. Our baseline estimate indicates that HSR connection raises co-publications by 35.2 percent. At the field level, effects are uniformly positive but range from 19.8 to 45.1 percent, and a one-collaborator increase in mean team size is associated with a 2.7 percentage-point larger HSR effect, with a correlation coefficient of 0.454 that is significant at the 5 percent level.

This paper makes two contributions to the literature on transportation, innovation, and knowledge production. First, our use of OpenAlex publication data combined with a route-distance-based definition of HSR connection produces a substantially broader empirical picture of HSR's effect on Chinese scientific collaboration than has previously been available. Existing publication-based work in this area, most notably Dong et al. (2020), relies on the Web of Science, which is weighted toward internationally visible, high-impact journals. OpenAlex, by contrast, provides materially wider coverage of Chinese research output, including domestic journals and lower-tier outlets, so that our estimates speak to a more complete spectrum of scientific collaboration. Second, and more substantively, we are - to our knowledge - the first to document that the aggregate HSR collaboration effect masks systematic heterogeneity across scientific fields, and that this heterogeneity is explained by field complexity. This links the literature on transportation and collaboration to the literature on economic complexity, and clarifies the channel through which HSR affects the composition of knowledge production rather than only its aggregate level.

The remainder of the paper is organised as follows. Section 2 reviews related literature. Section 3 describes the data, the construction of the HSR connection variable, and our measure of field complexity. Section 4 presents the empirical strategy. Section 5 reports the baseline and field-level results. Section 6 discusses the findings and their limitations. Section 7 concludes.

motivate this choice in Section 3 and discuss its limitations in Section 6.

2 Related Literature

Transportation infrastructure has long been associated with productivity and economic growth (Aschauer 1989; Munnell 1990), and more recent work emphasises its role in facilitating the knowledge flows that underpin regional innovation (Agrawal et al. 2017). Within this broader literature, Chinese HSR has attracted particular attention. Chen and Haynes (2017) show that HSR has contributed to regional economic convergence; Li et al. (2020) document positive effects on urban economic efficiency; and Tian et al. (2021) find that HSR promotes the agglomeration of service industries in peripheral cities. Together these studies establish HSR as a first-order shock to the economic geography of China.

A second strand of work examines the relationship between HSR and innovation, with broadly consistent findings that HSR connectivity raises innovative output. Gao and Zheng (2020) test the companion innovation hypothesis and show that HSR stimulates firm-level innovation by enlarging market scale and diversifying innovation sources. At the city level, Wang and Cai (2020) find that HSR construction enhances urban innovation capacity through knowledge spillover channels, and studies using patent counts and citations document positive effects on both the quantity and the quality of innovation (Zheng et al. 2022; Hanley et al. 2021). Komikado et al. (2021) provide corroborating evidence from Japan, where HSR is positively associated with regional patent applications in regions with previously poor accessibility. A growing number of studies have extended this analysis to the city-pair level in order to isolate HSR's role in fostering cross-city collaboration. Yao and Li (2022) and Wang et al. (2022) find, using Chinese co-patent data, that HSR connections stimulate inter-city technological collaboration; Cui et al. (2020) document similar gains in university–industry co-patenting; and Kang et al. (2023) show that innovation benefits extend to indirectly connected city pairs. Dong et al. (2020) report analogous results for scientific co-publication, further documenting rising co-author productivity and the emergence of new co-authorship relationships following the connection of secondary cities to major centres.

More recently, a subset of this literature has turned to the structural composition of

HSR-induced innovation rather than its quantity alone. Chen and Guo (2023), using the Shannon–Wiener diversity index, find that HSR promotes patent diversity across industries. Dong et al. (2025), applying standard complexity indices from the economic complexity literature, show that HSR enables cities to diversify into less related and more complex technological domains. Sun et al. (2023) distinguish types of innovation at the city level and find that collaborative patents and patents in technologically complex industries benefit more from HSR connection than independent patents and patents in discrete industries, suggesting that HSR's innovation effects are concentrated in activities requiring greater coordination and knowledge diversity. Despite this progress, the question of whether HSR-induced collaborative gains themselves vary systematically across scientific fields of differing complexity has not been directly examined. The present paper takes up this question.

Building on this body of evidence, we test two hypotheses. The first follows directly from the existing literature on HSR and knowledge flows: by lowering the time cost of face-to-face interaction between researchers in connected cities, HSR should raise the rate of scientific co-publication between city-pairs that gain a direct connection. Although video conferencing and digital communication have reduced some of the frictions associated with long-distance collaboration, a substantial body of work documents that in-person interaction remains important for building trust, transferring tacit knowledge, and coordinating complex joint work (e.g., Catalini 2018; Boudreau et al. 2017). We therefore expect HSR connections to generate a positive and economically meaningful increase in co-publication activity between linked city-pairs, relative to otherwise comparable unconnected pairs.

Hypothesis 1. *HSR connection between two cities increases scientific co-publication activity between them.*

Our second hypothesis concerns heterogeneity in this effect across scientific fields. If HSR facilitates collaboration primarily by lowering the cost of in-person coordination, then the size of its effect in any given field should depend on how intensively that field relies on coordination among multiple researchers. Fields characterised by a finer division of cognitive labour - where producing a single publication typically requires assembling complementary expertise from several researchers - stand to benefit more from a reduction in meeting costs than fields where research is predominantly carried out by solo authors or small, co-located teams. We follow Jones (2009) and a subsequent literature on the burden of knowledge in using the average number of authors per publication as a field-level proxy for the depth of knowledge specialisation and the need for coordinated teamwork (Balland et al., 2020): fields in which frontier research increasingly requires large teams are fields in which the marginal collaborator is more costly to reach and therefore more valuable to connect. This interpretation is consistent with the evidence, reviewed above, that HSR-induced innovation gains are concentrated in technologically complex and coordination-intensive domains (Sun et al. 2023; Dong et al. 2025). If the same mechanism operates for scientific collaboration, the HSR treatment effect should be systematically larger in more team-intensive fields.

Hypothesis 2. *The effect of HSR connection on scientific co-publication is increasing in the average number of authors per publication in a field, a proxy for the complexity and division of knowledge required to produce research in that field.*

3 Data and Variables

3.1 Publication and HSR Data

Our analysis draws on two principal data sources. Publication data come from OpenAlex, from which we extract the universe of scholarly publications with at least one co-author affiliated with a Chinese research institution over the period 2000 to 2025. For each publication, we record the research field, the number of co-authors, and the city of each co-author's host institution. Co-publication counts at the city-pair–year level are constructed as follows: a publication contributes one unit to the count of city pair (i, j) in year t if and only if at least one co-author is affiliated with a city-i institution and at least one co-author with a city-j

institution. Multiple co-authors from the same city are not double-counted, so that the measure captures the existence of an inter-city collaborative tie rather than the size of the team. We denote the resulting city-pair–year co-publication count by $Counts_{p,t}$. Descriptive statistics are reported in Table 1, Panel A.

Table 1: Descriptive Statistics

Variable	Explanation	Obs.	Mean	Std. dev.	Min	Max
Panel A: city-pair, year, field						
Year	Publication year	1,285,196	2,018.45	5.46	2,000	2,025
Counts	Number of co-publications	1,285,196	11.53	75.36	1	8,712
Connected	Whether city-pair is HSR connected	1,285,196	0.59	0.49	0	1
connect_year	Year of HSR connection (0 if never)	970,227	1,782.68	643.75	0	2,025
geo_distance	Geographical distance (km)	970,227	1,016.55	693.51	16.29	4,800.43
route_distance	Shortest HSR route distance (km)	917,696	1,348.04	853.53	20.55	6,489.08
Panel B: field						
Collaborators	Average number of collaborators	20	6.12	1.30	3.57	8.93

High-speed rail data are drawn from the National Railway Administration, the Railway 12306 platform, and local news sources.[3] Since the first HSR line opened in 2008, the network has expanded rapidly and, as of 2020, covers 270 prefecture-level cities in our dataset. For each station that came into operation along a newly constructed HSR line during our sample period, we record the host city, province, and opening year. We restrict attention to stations offering passenger service. Summary statistics for the HSR data are reported in Table 1, Panel A.

3.2 Construction of the Treatment Variable

We construct the treatment variable from the HSR network as follows. In each year of the sample we compute the length of the shortest HSR route between every pair of cities. A city pair p is classified as HSR-connected in year t if the shortest feasible HSR route between the

[3] We thank Zhaoyingzi Dong of Zhejiang University for generously sharing the HSR data. For further details, see Dong et al. (2025).

two cities is no longer than twice the great-circle geographical distance between them. This rule captures direct and short multi-segment HSR connections while excluding detour-heavy routes that would not plausibly substitute for conventional travel. For each city we use the mean latitude and longitude of its research institutions as a proxy for city location; limitations of this approximation are discussed in Section 6. Formally, the treatment indicator is defined as

$$Connected_{p,t} = \begin{cases} 1, & \text{if } p \text{ is connected at } t \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Of the 33,793 city pairs observed across the sample period, 17,558 become HSR-connected at some point between 2008 and 2020 (Table 1, Panel A).

3.3 Field-Level Sample and Complexity Proxy

In the second half of the analysis we decompose the panel by scientific field and examine whether the estimated HSR treatment effect varies systematically across fields with differing levels of complexity. We adopt the twenty scientific fields defined by OpenAlex, which are grouped into four umbrella domains - Physical Sciences, Health Sciences, Life Sciences, and Social Sciences. Our primary analysis covers the first three; we motivate the exclusion of Social Sciences below. Fields are the second-least-granular level of the OpenAlex taxonomy, below which subfields and topics are nested in increasing specificity. Field-level co-publication counts, $Counts^{f}_{p,t}$, are constructed analogously to the aggregate measure.

As discussed above, we proxy the relative complexity of each field by the mean number of co-authors per publication within the field, which we denote $Collaborators^{f}$. Average team size is not perfect but maps onto the theoretical construct we seek to capture here. A larger average team reflects a deeper division of cognitive labour, which is itself a defining feature of complex knowledge production (Balland et al. 2020). The team-science literature provides independent support: Jones (2009) argues that the expansion of the knowledge frontier compels individual researchers to narrow their expertise, which in turn raises the optimal team size required to assemble the combination of specialisations needed for any given research problem. Wuchty, Jones, and Uzzi (2007) document that team-authored research has grown across essentially all scientific fields and that this growth has been most pronounced in fields requiring integration across specialised knowledge bases. Consistent with this logic, the fields with the largest mean team sizes in our sample - Immunology and Microbiology (8.9 co-authors on average), Biochemistry, Genetics and Molecular Biology (8.0), and Medicine (7.7) - are those widely understood to require laboratory infrastructure, multi-method experimental coordination, and cross-disciplinary expertise. The smallest mean team sizes appear in Mathematics (3.6) and Computer Science (3.7), fields in which individual or small-team theoretical work remains common. This ordering is broadly consistent with what one would expect if team size reflects the depth of knowledge specialisation and the associated coordination burden. We nevertheless acknowledge that team size can also be shaped by factors unrelated to complexity, including funding structures, data-sharing norms, and disciplinary conventions regarding authorship; we return to this caveat in Section 6. Summary statistics for $Collaborators^f$ are reported in Table 1, Panel B.

We restrict the field-level sample to fields in the Physical, Health, and Life Sciences, excluding Social Sciences, primarily because pooling these fields would likely bias our cross-field test by potentially introducing variation in team size that might reflect disciplinary convention rather than the coordination requirements we aim to measure.

4 Empirical Strategy

We estimate the causal effect of HSR connection on scientific co-publication using a staggered difference-in-differences design. The baseline specification is

$$\log(Counts_{p,t}) = \beta\, Connected_{p,t} + \alpha_p + \lambda_t + \varepsilon_{p,t}, \tag{2}$$

where β is the coefficient of interest, α_p denotes city-pair fixed effects, λ_t denotes year fixed effects, and $\varepsilon_{p,t}$ is an idiosyncratic error term. The city-pair fixed effects absorb time-invariant determinants of collaborative propensity between any two cities - including their geographical distance, institutional composition, historical research ties, and linguistic or cultural affinities - while the year fixed effects absorb common shocks to Chinese scientific productivity such as aggregate funding cycles or nationwide policy changes.

Because the outcome $\log(Counts_{p,t})$ is undefined when $Counts_{p,t} = 0$, equation (2) is estimated on the subsample of city-pair-years with at least one co-publication (i.e., $Counts_{p,t} \geq 1$).

Because HSR openings occur at different dates across city pairs and because the treatment effect may be heterogeneous across cohorts, the two-way fixed effects estimator is known to be biased in this setting (Goodman-Bacon 2021; de Chaisemartin and D'Haultfœuille 2020). We therefore estimate β using the multiple-period difference-in-differences framework of Callaway and Sant'Anna (2021), which recovers group–time average treatment effects on the treated, $ATT_{g,t}$, by comparing each cohort of city pairs first connected in year g to a control group of not-yet-treated city pairs in each post-treatment year t. This estimator is robust to staggered adoption and to heterogeneous treatment effects across cohorts and horizons. We implement it using the Stata package `CSDID` (Rios-Avila et al. 2021), and aggregate the $ATT_{g,t}$ into a single summary parameter θ for the average effect of HSR connection on city-pair co-publication.[4] An event-study decomposition of the $ATT_{g,t}$ across event time is reported in Figure 1 and is used below to assess the plausibility of the parallel-trends assumption.

To examine how HSR treatment effects vary across scientific fields, we re-estimate the baseline specification separately for each of the twenty OpenAlex fields:

$$\log\big(Counts^f_{p,t}\big) = \beta^f \, Connected_{p,t} + \alpha^f_p + \lambda^f_t + \varepsilon^f_{p,t}, \tag{3}$$

[4]For details of the estimator, see Callaway and Sant'Anna (2021).

where α_p^f and λ_t^f are field-specific city-pair and year fixed effects, and β^f is the field-specific coefficient of interest. For each field f we aggregate the estimated group–time effects into a single field-level treatment parameter ATT^f. We then relate these field-level effects to our complexity proxy by estimating the cross-field regression

$$ATT^f = \beta\, Collaborators^f + \alpha + \varepsilon^f, \tag{4}$$

where β is the coefficient of interest, α is a constant, and ε^f is the error term. This regression has twenty observations - one per field - and is therefore best interpreted as a descriptive test of whether the cross-field dispersion of HSR treatment effects lines up with cross-field variation in team size. We return to the inferential limits of this specification in Section 6.

5 Results

5.1 Baseline Effect on Scientific Co-publication

HSR connection produces a large and statistically significant increase in scientific co-publication at the city-pair level. The aggregate ATT reported in Table 2, Panel A is 0.301 (standard error 0.013), significant at the 1 percent level. Exponentiating yields a 35.2 percent increase in co-publication counts following HSR connection. Given that the average city pair in our sample produces approximately 11.5 co-publications per year in any given field (Table 1, Panel A), this effect is economically substantial. The estimate indicates that reducing the cost of inter-city travel through fixed-route transportation infrastructure meaningfully raises the propensity for cross-city scientific collaboration.

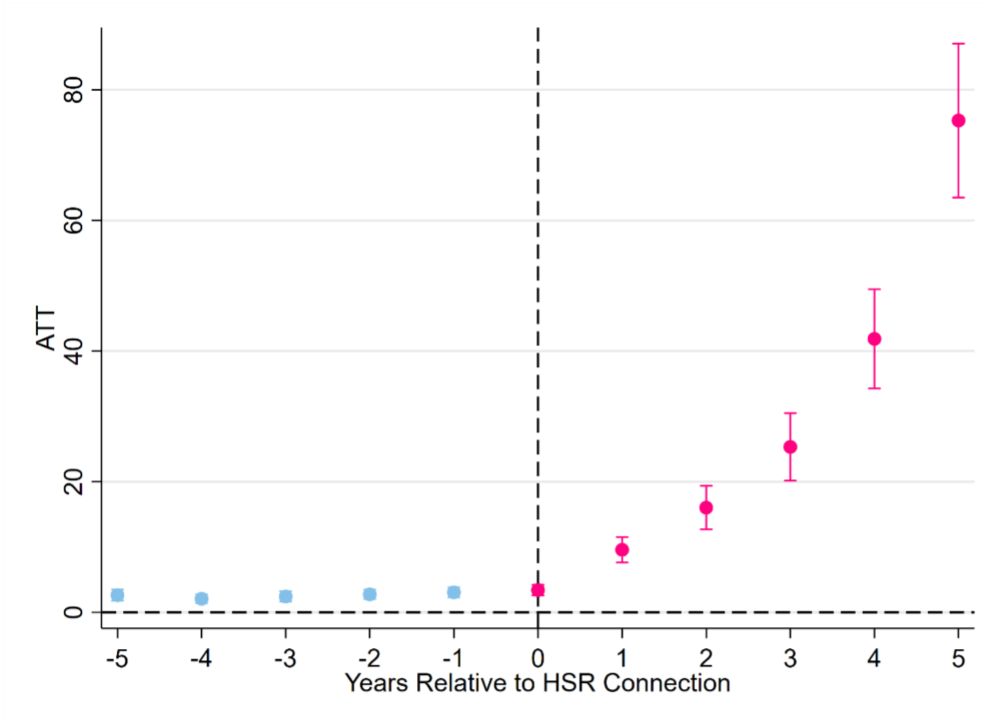

Figure 1: HSR connection on co-publication event-study time graph

Figure 1 reports the event-study decomposition of the ATT for five years before and five years after HSR connection. The pre-treatment coefficients display no systematic upward or downward trend: they lie in a narrow band and are economically indistinguishable from one another, consistent with the parallel-trends assumption underlying identification. Although the pre-treatment estimates are not statistically zero, this reflects the size of the panel (970,227 city pairs, reported in Table 1, Panel A) rather than any substantive pre-trend, as the coefficients are roughly an order of magnitude smaller than the post-treatment estimates. Following HSR connection, the treatment effect becomes positive and grows monotonically over the post-treatment horizon, consistent with a dynamic process in which new connections first enable initial inter-city interactions among researchers, which subsequently develop into sustained collaborative relationships that produce co-authored publications over time.

Table 2: Baseline and Field-level Results

Variable	Aggregate *ATT*	Std. err.	% change	95% CI	p-value
Panel A: baseline results					
log(*Counts*)	0.3015	0.0134	35.19	[0.2751, 0.3278]	< 0.001
Panel B: field-level results					
Agricultural and Biological Sciences	0.2922	0.0221	33.93	[0.2488, 0.3355]	< 0.001
Biochemistry, Genetics and Molecular Biology	0.3640	0.0178	43.91	[0.3292, 0.3988]	< 0.001
Chemical Engineering	0.3005	0.0435	35.06	[0.2152, 0.3859]	< 0.001
Chemistry	0.1945	0.0231	21.47	[0.1493, 0.2397]	< 0.001
Computer Science	0.2523	0.0227	28.70	[0.2079, 0.2967]	< 0.001
Dentistry	0.2790	0.0560	32.18	[0.1692, 0.3887]	< 0.001
Earth and Planetary Sciences	0.2974	0.0309	34.63	[0.2368, 0.3580]	< 0.001
Energy	0.3609	0.0387	43.46	[0.2850, 0.4368]	< 0.001
Engineering	0.2362	0.0173	26.64	[0.2024, 0.2700]	< 0.001
Environmental Science	0.3073	0.0219	35.97	[0.2643, 0.3503]	< 0.001
Health Professions	0.3088	0.0407	36.18	[0.2291, 0.3886]	< 0.001
Immunology and Microbiology	0.3344	0.0266	39.71	[0.2823, 0.3865]	< 0.001
Materials Science	0.2250	0.0204	25.23	[0.1849, 0.2650]	< 0.001
Mathematics	0.1805	0.0277	19.78	[0.1262, 0.2348]	< 0.001
Medicine	0.3594	0.0169	43.24	[0.3262, 0.3925]	< 0.001
Neuroscience	0.2420	0.0283	27.38	[0.1865, 0.2975]	< 0.001
Nursing	0.2881	0.0366	33.39	[0.2165, 0.3598]	< 0.001
Pharmacology, Toxicology and Pharmaceutics	0.2456	0.0324	27.84	[0.1821, 0.3092]	< 0.001
Physics and Astronomy	0.3725	0.0300	45.14	[0.3138, 0.4313]	< 0.001
Veterinary	0.2028	0.0479	22.48	[0.1090, 0.2967]	< 0.001

5.2 Heterogeneity across Scientific Fields

Table 2, Panel B reports field-specific treatment effects for each of the twenty OpenAlex fields. All twenty point estimates are positive and statistically significant at the 1 percent level, indicating that HSR connection promotes co-publication broadly across the sciences. The magnitude of the effect, however, varies considerably across fields. Physics and Astronomy exhibits the largest estimated ATT at 0.373 (a 45.1 percent increase in co-publication), followed closely by Biochemistry, Genetics and Molecular Biology and by Energy. At the lower end, Mathematics displays the smallest effect at 0.181 (19.8 percent), followed by Chemistry and Veterinary Science. The spread between the highest and lowest field-level ATT is approximately 25 percentage points. Figure 2 plots the field-level $ATTs$ in descending order and confirms that the aggregate baseline result conceals substantial cross-field heterogeneity.

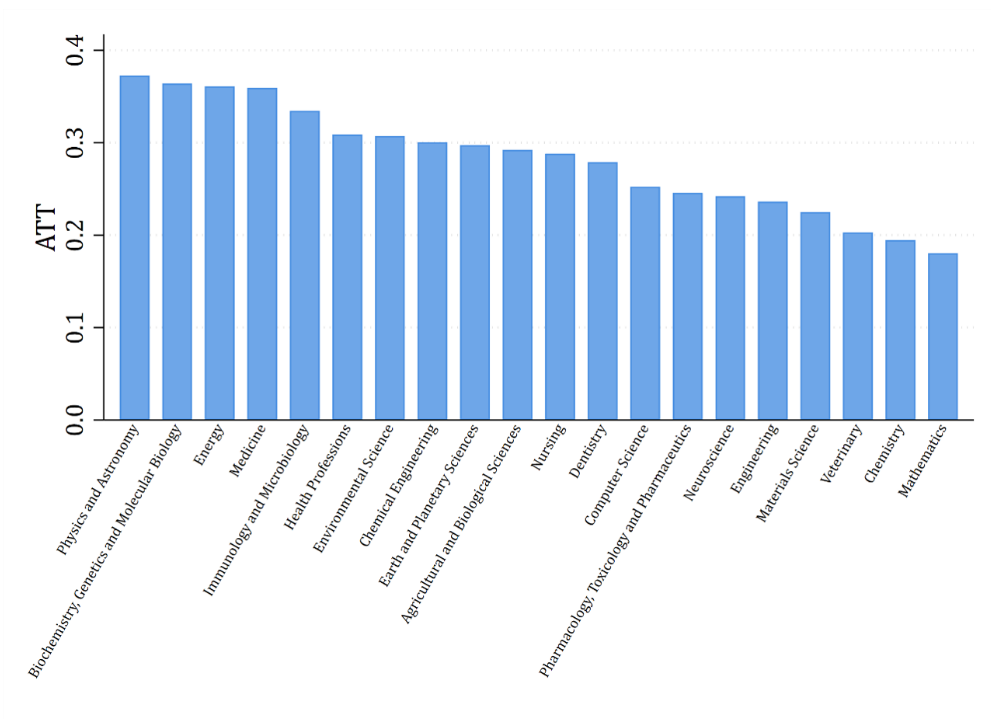

Figure 2: Field-level HSR effects for scientific fields

Table 3: Average Team Size on HSR Effect

Dependent variable	Regressor	Coefficient	Std. err.	p-value	Obs.	R^2
Aggregate ATT	*Collaborators*	0.0205	0.0095	0.044	20	0.206

Correlation coefficient: 0.454.

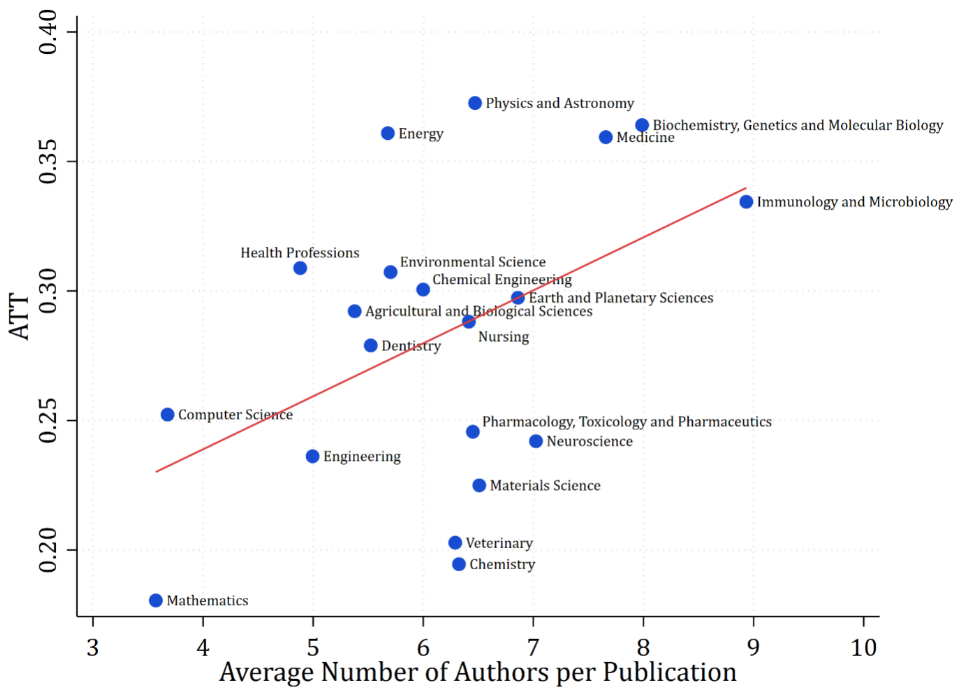

Figure 3: HSR effect and average team size with regression line for scientific fields

Figure 3 plots each field's estimated ATT against its average number of collaborators per publication, together with the fitted OLS regression line. The regression, reported in Table 3, yields a slope coefficient of 0.0205 (standard error 0.0095, $p = 0.044$) and a correlation coefficient of 0.454, significant at the 5 percent level. Interpreted in percentage terms, a one-collaborator increase in mean team size is associated with an approximately 2.7 percentage-point larger HSR treatment effect. The R^2 of 0.21 indicates that average team size alone accounts for roughly one-fifth of the cross-field variation in estimated HSR effects. The geometry of Figure 3 matches this quantitative summary: Physics and Astronomy, Biochemistry, and Medicine cluster in the upper portion of the plot, combining high average team sizes with large HSR effects, while Mathematics and Computer Science occupy the lower-left corner with small team sizes and small effects. These results are consistent with the theoretical prediction that reductions in the cost of face-to-face interaction should disproportionately benefit knowledge production that relies on deeper divisions of specialised cognitive labour.

6 Discussion

Our baseline estimates indicate that HSR connection generates a substantial increase in cross-city scientific collaboration in China. The most directly comparable prior estimate is Dong et al. (2020), who report an 11 percent increase in co-publication following HSR connection. Our estimate of 35.2 percent is materially larger. We attribute the gap to three differences in design. First, we use OpenAlex, which provides far broader coverage of Chinese research output than the Web of Science database used by Dong et al. (2020); OpenAlex includes domestic journals and lower-tier outlets to which HSR-induced collaboration is likely to be particularly responsive. Second, our sample period (2000–2025) is longer than the 2006–2016 window used by Dong et al. (2020), allowing us to observe longer post-treatment horizons during which collaborative ties mature into co-publications. Third, our route-distance-based definition of HSR connection captures a broader set of connected pairs than the direct-connection definition used in the prior work. Read in combination, the larger effect and broader empirical footprint are consistent with HSR's collaborative benefits being concentrated among smaller and more peripheral cities - a pattern also anticipated by Dong et al. (2020) but not fully quantified there.

The field-level results support the view that the collaborative returns to HSR depend on the structure of the underlying knowledge. More complex economic activities require the coordination of larger numbers of increasingly specialised individuals (Balland et al. 2020), and infrastructure that reduces the cost of face-to-face interaction should therefore generate disproportionately large collaborative gains where such coordination is most demanding (Agrawal et al. 2017). The closest antecedent in the HSR literature is Dong et al. (2025), who apply standard complexity indices to patent data and show that HSR enables cities to diversify into less related and more complex technological fields. Our analysis is complementary: where Dong et al. (2025) focus on how HSR alters the composition of a city's technology portfolio, we identify a mechanism upstream of that composition by showing that the co-production of knowledge itself responds to HSR in a complexity-graded fashion.

The cross-field pattern also offers internal support for our decision to exclude the Social Sciences from the complexity analysis. In our data, the mean number of collaborators per publication across the Life, Health, Physical, and Social Science domains is 7.12, 6.19, 5.58, and 3.08 respectively. The Social Sciences fall far below the other three, by a margin that is difficult to reconcile with a pure complexity interpretation. Rather, the discrepancy is consistent with the Social Sciences operating under fundamentally different authorship and team-formation conventions, as documented by Wuchty et al. (2007). This reinforces the concern that average team size does not map onto complexity in a comparable way across these domains, and motivates our restriction to the natural and applied sciences for the complexity test.

Several limitations warrant acknowledgement. First, our HSR network is constructed using the mean latitude and longitude of research institutions as a proxy for city location. Under this proxy, the city of Yinchuan can never be classified as HSR-connected, and we therefore exclude all Yinchuan-containing pairs from both treatment and control groups. Yinchuan represents a small share of the panel, and this exclusion should not materially affect our estimates; a more complete resolution would require comprehensive city-level geographical coverage, which we leave to future work. Second, our analysis is confined to China. China's HSR network is unusually extensive and its research system is shaped by centralised funding and institutional structures that differ from those of other large research-producing economies; the external validity of our estimates to other contexts is therefore an open question. Third, although the baseline treatment effect is identified within a difference-in-differences design that supports a causal interpretation, the cross-field relationship between HSR effects and team size is estimated on twenty observations and is inherently correlational. It is sensitive to individual fields and to the choice of complexity proxy. As discussed in Section 3, average team size is an approximate measure of complexity that may also reflect funding conventions, data-sharing norms, and authorship practices that vary across fields independently of the coordination demands of the science. Fourth, although we interpret our findings through the lens of face-to-face interaction - in line with the broader literature on transportation and knowledge flows - HSR may also affect collaboration through other channels, including permanent researcher migration and institutional hiring decisions. Disentangling these mechanisms is a natural direction for future work.

7 Conclusion

This paper studies the effect of China's high-speed rail expansion on scientific collaboration and asks whether the collaborative returns to HSR are systematically larger for more complex fields of science. Using a staggered difference-in-differences design applied to city-pair–year co-publication data from OpenAlex, we estimate that HSR connection raises cross-city scientific co-publication by 35.2 percent. At the field level, treatment effects are uniformly positive but vary materially across fields, and this variation is positively and significantly correlated with field complexity as proxied by average team size. Together, these findings indicate that transportation infrastructure not only stimulates scientific collaboration broadly but does so disproportionately for fields that rely on deep divisions of specialised knowledge.

The results carry implications along two dimensions. First, from a policy perspective, they suggest that HSR is not a neutral instrument with respect to the composition of scientific output. The collaborative returns are unevenly distributed across fields, with the largest gains accruing to activities that depend on coordination among highly specialised researchers.

Infrastructure planners and innovation-policy designers evaluating prospective HSR corridors on the basis of their expected scientific returns should therefore attend not only to the density of research institutions in candidate cities but also to the field composition of those institutions, given that the marginal returns to connection appear greatest where complex, collaboration-intensive science is already present. Second, from a scholarly perspective, this paper contributes to the literatures on transportation, innovation, and economic complexity by documenting, for the first time to our knowledge, that the aggregate effect of HSR on scientific collaboration conceals systematic heterogeneity explained by field complexity. The use of OpenAlex rather than the Web of Science also broadens the empirical scope of the HSR–collaboration literature, indicating that previously documented effects generalise to a wider slice of the Chinese research ecosystem than the internationally visible segment alone.

Several avenues remain open for future research. Extending the analysis to other national HSR systems would establish whether the complexity–collaboration relationship documented here is a general property of high-speed transportation or a feature specific to the Chinese research landscape. Replication using alternative data sources - including patent data, clinical-trial registries, and grant-level records - would help discipline the team-size proxy used here and shed light on the mechanisms through which face-to-face interaction shapes the production of complex knowledge.

References

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

Aschauer, D.A., 1989. Is public expenditure productive? *Journal of Monetary Economics* 23 (2), 177–200.

Balland, P.-A., Boschma, R., Crespo, J., Rigby, D.L., 2019. Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies* 53 (9), 1252–1268.

Balland, P.-A., Jara-Figueroa, C., Petralia, S.G., Steijn, M.P.A., Rigby, D.L., Hidalgo, C.A., 2020. Complex economic activities concentrate in large cities. *Nature Human Behaviour* 4 (3), 248–254.

Balland, P.-A., Rigby, D., 2017. The geography of complex knowledge. *Economic Geography* 93 (1), 1–23.

Becker, G.S., Murphy, K.M., 1992. The division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics* 107 (4), 1137–1160.

Callaway, B., Sant'Anna, P.H.C., 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics* 225 (2), 200–230.

Chen, Z., Guo, X., 2023. High-speed rail and industrial technological diversification: evidence from China. [Journal], [vol].

Chen, Z., Haynes, K.E., 2017. Impact of high-speed rail on regional economic disparity in China. *Journal of Transport Geography* 65, 80–91.

Cui, L., Chen, Z., Jiang, H., Zhang, Y., 2020. University-industry co-invention and high-speed rail: evidence from China.

de Chaisemartin, C., D'Haultfœuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110 (9), 2964–2996.

Dong, X., Zheng, S., Kahn, M.E., 2020. The role of transportation speed in facilitating high-skilled teamwork across cities. *Journal of Urban Economics* 115, 103212.

Dong, Z., Balland, P.-A., et al., 2025. High-speed rail and technological diversification: evidence from China.

Gao, Y., Zheng, J., 2020. The impact of high-speed rail on innovation: firm-level evidence from China.

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

Hanley, D., Li, J., Wu, M., 2021. High-speed rail and innovation in China.

Hidalgo, C.A., Hausmann, R., 2009. The building blocks of economic complexity. *PNAS* 106 (26), 10570–10575.

Jones, B.F., 2009. The burden of knowledge and the "death of the Renaissance man": is innovation getting harder? *Review of Economic Studies* 76 (1), 283–317.

Kang, L., et al., 2023. Indirect high-speed rail connection and urban innovation: evidence from China.

Komikado, H., Morikawa, S., Bhattacharya, A., Kato, H., 2021. High-speed rail, inter-regional accessibility, and regional innovation: evidence from Japan. *Technological Forecasting and Social Change* 167, 120697.

Li, P., Rigby, D., 2022. Relatedness, complexity, and economic growth in Chinese cities. *International Regional Science Review*.

Li, Y., Chen, Z., Wang, P., 2020. High-speed rail and urban economic efficiency: evidence from China.

Munnell, A.H., 1990. Why has productivity growth declined? Productivity and public investment. *New England Economic Review*, 3–22.

Rios-Avila, F., Callaway, B., Sant'Anna, P.H.C., 2021. CSDID: Stata module for the estimation of difference-in-differences with multiple time periods.

Sun, Y., et al., 2023. High-speed rail, innovation types, and technological complexity.

Tian, G., Li, Y., Jiang, Y., 2021. High-speed rail and service industry agglomeration in Chinese peripheral cities.

Wang, F., Wei, X., Liu, J., He, L., Gao, T., 2022. Impact of high-speed rail on inter-city technological collaboration: evidence from Chinese co-patents.

Wang, L., Cai, W., 2020. High-speed rail and urban innovation: knowledge spillover channels. *Regional Studies*.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. *Science* 316 (5827), 1036–1039.

Yao, S., Li, Z., 2022. High-speed rail and cross-city innovation collaboration in China.

Zhang, L., et al., 2024. OpenAlex coverage of Chinese research output: a comparative assessment.

Zheng, S., Du, R., Wang, P., 2022. High-speed rail and urban innovation: evidence from patents and citations.