Available Online: 14 September 2026 DOI: 10.54254/2977-5701/2026.36794

First access and network densification: high-speed rail expansion and urban industrial structure in China

Xuyuan Tao

School of Economics, University of International Business and Economics, Beijing, China

202402055@uibe.edu.cn

Abstract. This paper asks whether the structural consequences of High-Speed Rail (HSR) expansion arise from a city's first connection to the network or from subsequent network densification. This paper matches a hand-coded database of passenger-rail line openings to a panel of Chinese prefecture-level cities from 2008 to 2019 and exploits within-city changes in access in a two-way fixed effects framework. First HSR access increases the tertiary-sector share of Gross Domestic Product (GDP) by 0.925 percentage points and reduces the secondary-sector share by 1.078 percentage points. The reallocation is most pronounced in western cities. By contrast, once initial access is accounted for, additional lines and hub status yield statistically imprecise average effects. A pretreatment placebo and a 2014-cohort difference-in-differences design corroborate the timing and direction of the baseline estimates. These findings identify network entry, rather than subsequent densification, as the principal margin through which HSR expansion reshapes urban industrial structure. They imply that the structural returns to extending network coverage can differ materially from those of adding capacity and connectivity within already-served cities.

Keywords: high-speed rail, industrial structure, extensive margin, regional heterogeneity

---

1. Introduction

As China's High-Speed Rail (HSR) network matures, policy attention is shifting from connecting previously unserved cities to deepening connectivity within the existing network. Whether first access and subsequent network expansion exert different effects on urban sectoral composition is therefore both an important empirical question and a consequential policy question.

Donaldson and Hornbeck showed that rail access substantially affected nineteenth-century U.S. agricultural land values [1]. Faber found that highway connections to larger cities reduced Gross Domestic Product (GDP) growth in peripheral Chinese counties [2]. Banerjee et al. reported positive level effects but no effect on per capita GDP growth [3]. Baum-Snow et al. showed that network configuration also shapes industrial location [4]. Existing studies link HSR access to changes in industrial composition and propose innovation, knowledge spillovers, and producer-service agglomeration as potential channels [5-8]. However, most studies use a binary HSR indicator, leaving the distinction between first access and subsequent network expansion less clear. Transport improvements may generate either agglomeration or displacement depending on location and prior connectivity. Whether becoming connected for the first time differs from gaining additional lines or hub status therefore remains an important empirical question.

This paper uses a hand-collected dataset of designated high-speed and passenger-dedicated rail line segments mapped to the prefecture-year level. A city is classified as connected when at least one included segment is operating and serves a station mapped to that city. First access changes a city's network status, whereas additional lines deepen connectivity conditional on existing access. A two-way fixed effects model, regional interaction estimates, and a 2014-cohort difference-in-differences comparison estimate how these two margins affect urban industrial composition.

This study contributes in three ways. First, it builds a city-year measure separating initial HSR access from later line additions and hub formation. Second, it compares these margins within a common framework and reports formal coefficient-difference tests. Third, it evaluates regional heterogeneity and compares the relative robustness of tertiary- and secondary-sector responses. Together, these analyses clarify how the implications of HSR expansion differ between extending network coverage and densifying an existing network.

2. Data, variables, and econometric model

2.1. Data sources

City-year data for 2008-2019 are compiled from annual editions of the China City Statistical Yearbook and include three-sector GDP shares, population, fiscal expenditure, Foreign Direct Investment (FDI), urban construction land, and road area; the annual tables were downloaded from the China National Knowledge Infrastructure (CNKI) Yearbook Database [9]. The HSR measure is constructed from a hand-collected list of designated passenger-rail line segments, with line inclusion guided by the "Eight Vertical, Eight Horizontal" framework in the *2016 Medium- and Long-Term Railway Network Plan* [10]. Operating dates and station locations are mapped to prefecture-level cities, and the line segments serving each city are counted. All HSR regressors are lagged one year, so baseline estimation begins in 2009. The measure captures both first access and subsequent network expansion instead of relying exclusively on a binary indicator or selected train-service categories. Segmented openings and differences between planning corridors and physical line segments may nevertheless introduce measurement error. The sample ends in 2019 to avoid FDI data discontinuities and major external disruptions from 2020 onward. The unbalanced raw panel contains 3,475 observations for 297 city codes; the complete-case specification includes 2,564 observations for 282 cities. Controls include economic fundamentals (ln pgdp, ln pop), institutional and openness indicators (govrate and fixed-rate FDI intensity), and urban infrastructure (ln urban, ln road).

2.2. Variable definitions

The dependent variables are the GDP shares of the tertiary, secondary, and primary sectors (%). hsrdummy equals 1 if at least one included HSR line segment operates in a city; hsrlines counts operating line segments; and hub identifies cities served by at least three lines. All HSR regressors are lagged one year. The intensive-margin variable, L.addlines, equals L.hsrlines minus L.hsrdummy, with missing lagged observations preserved as missing. The observed lagged line count ranges from 0 to 7. Controls comprise ln pgdp, ln pop, govrate, fdi_rate_fixed2019, ln urban, and ln road. FDI intensity converts reported U.S.-dollar FDI to Renminbi (RMB) at the fixed 2019 annual-average rate of 6.8985 RMB per U.S. dollar before normalization by GDP. The conversion standardizes the currency denomination while preserving the original series' cross-city and intertemporal variation. Table 1 reports descriptive statistics.

Table 1. Descriptive statistics

Variable	Obs.	Mean	S.D.	Min	Max
Tertiary share (%)	2,564	40.73	10.09	11.47	83.52
Secondary share (%)	2,564	47.42	10.25	10.68	87.96
Primary share (%)	2,564	11.85	7.78	0.03	49.89
L.hsrdummy	2,564	0.39	0.49	0.00	1.00
L.hsrlines	2,564	0.58	0.92	0.00	7.00
L.hub	2,564	0.04	0.20	0.00	1.00
ln(GDP per capita)	2,564	10.62	0.60	8.70	12.46
ln(Population)	2,564	5.94	0.64	3.43	8.14
Gov. expenditure / GDP (%)	2,564	18.85	9.24	4.39	101.75
FDI / GDP, fixed-rate (%)	2,564	1.87	1.85	0.00	13.41
ln(Urban construction land)	2,564	4.52	0.84	1.79	7.56
ln(Road area)	2,564	7.08	0.95	2.64	10.01

Note: HSR variables are lagged one year. Descriptive statistics correspond to the 2,564-observation complete-case baseline sample, which spans 2009-2019. Obs. = Observations; S.D. = Standard Deviation; Gov. = Government.

2.3. Econometric model

Following Wooldridge's panel-data framework, the baseline two-way fixed-effects specification is shown in Equation (1) [11].

$$Y_{it} = \beta \cdot L.hsr_{it} + \gamma X_{it} + \alpha_i + \lambda_t + \varepsilon_{it}(1) \tag{1}$$

where i indexes cities and t indexes years; Y is the sectoral GDP share; $L.hsr$ is the lagged HSR variable, represented by hsrdummy in the baseline and extended to additional lines and hub status; X is the vector of controls; and α_i and λ_t are city and year fixed effects. Under parallel sectoral trends across cities with different access timing, β identifies the effect of HSR access. City fixed effects absorb time-invariant local characteristics, year fixed effects capture common annual shocks, and standard errors are clustered at the city level. Because route placement may respond to time-varying prospects and some controls may respond to earlier HSR access, Section 5 examines pretreatment timing, alternative fixed effects, a cohort comparison, and same-sample sensitivity. The full-control coefficient is interpreted as a conditional treatment effect rather than an unconditional total effect.

3. Baseline results

Table 2 reports stepwise regressions for the tertiary and secondary sectors. All columns include city and year Fixed Effects (FE) with city-clustered standard errors. Column (1) includes no additional controls; column (2) adds economic fundamentals (ln pgdp, ln pop); column (3) adds the institutional and openness indicators govrate and fixed-rate FDI intensity; and column (4) adds urban infrastructure (ln urban, ln road). Missing control values cause the estimation sample to vary across these columns.

In the full specification, first HSR access raises the tertiary share by 0.925 percentage points and reduces the secondary share by 1.078 percentage points. Both effects are significant at the 1 percent level. The 0.925-

point tertiary increase equals approximately 2.3 percent of the complete-sample mean, as does the absolute 1.078-point secondary decrease, implying an economically meaningful shift in sectoral composition. Across the four columns, tertiary coefficients range from 0.815 to 0.925 and secondary coefficients from -0.922 to -1.078. Within R^2 rises from 0.698 to 0.757 for tertiary and from 0.477 to 0.634 for secondary. Because sample composition also changes, Section 5.4 reports same-sample comparisons to isolate coefficient movement from changes in observations.

The control variables are included to account for observed economic conditions, openness, public expenditure, and urban infrastructure. Their coefficients are not interpreted as separately identified policy effects because several may be jointly determined with HSR access. The primary object of interest is the stability and magnitude of the HSR coefficient across the stepwise specifications.

Table 2. Baseline regressions: stepwise inclusion of controls

Variable	Tertiary (1)	Tertiary (2)	Tertiary (3)	Tertiary (4)	Secondary (1)	Secondary (2)	Secondary (3)	Secondary (4)
L.hsrdummy	0.827**	0.815**	0.836**	0.925***	-1.054**	-0.922***	-1.018***	-1.078***
	(0.387)	(0.339)	(0.331)	(0.326)	(0.480)	(0.356)	(0.349)	(0.352)
ln pgdp	—	-9.058***	-8.230***	-8.425***	—	16.035***	15.329***	15.513***
ln pop	—	-7.635***	-4.768**	-3.172*	—	3.105*	2.393	0.807
govrate	—	—	0.055**	0.050**	—	—	-0.039	-0.035
fdirate (fixed-rate)	—	—	0.302***	0.287***	—	—	-0.167	-0.134
ln urban	—	—	—	-0.400	—	—	—	0.602
ln road	—	—	—	0.484	—	—	—	-0.220
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City-clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Within R^2	0.698	0.737	0.744	0.757	0.477	0.617	0.626	0.634
N	3,164	3,146	2,683	2,564	3,164	3,146	2,683	2,564

Note: Parentheses report city-clustered standard errors for the HSR coefficient; significance markers for all coefficients use the same clustering. Control-variable coefficients are shown directly, while their standard errors are not separately displayed to preserve readability. Dashes indicate variables not included in a column. All controls enter as continuous variables. FE = Fixed Effects; SE = Standard Errors. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels respectively.

4. Heterogeneity and extended analysis

4.1. Regional heterogeneity

Eastern, central, and western regions differ in development levels and transport-infrastructure endowments, so national averages may conceal substantial heterogeneity. Using the historical three-region classification employed in official statistics, Table 3 reports both regional subsample regressions and a pooled regional-interaction model.

To compare regional effects directly in the pooled sample, this paper extends Equation (1) to the regional-interaction specification shown in Equation (2), with estimates reported in Panel B of Table 3. Here, $r \in \{E, C, W\}$ denotes East, Central, and West, $Region_{ir}$ is the corresponding regional indicator, and no separate HSR-access term is included:

$$Y_{it} = \sum_r \in \{E, C, W\}\beta_r(L.HSR_{it} \times Region_{ir}) + \gamma' X_{it} + \alpha_i + \lambda_t + \varepsilon_{it}(2) \quad (2)$$

The pooled estimates show positive tertiary-sector effects in the central and western regions but no precise effect in the east. Both the central and western effects exceed the eastern effect, while they do not differ from each other. For the secondary share, the estimates are negative in the central and western regions and imprecise in the east. The west-east difference is significant, the central-east difference is not, and the west-central difference is marginal. Panel A's separate-sample estimates further highlight the western pattern: first access raises the western tertiary share by 1.487 percentage points and reduces the secondary share by 2.145 points, both statistically significant. This pattern accords with economic-geography models in which transport costs, economies of scale, and market size shape core-periphery outcomes [12]. It indicates that HSR-induced structural adjustment is more pronounced where prior transport networks were less dense, although the data do not isolate a single channel.

Table 3. Regional heterogeneity: subsample estimates and direct regional effects

Variable	Tertiary: East	Tertiary: Central	Tertiary: West	Secondary: East	Secondary: Central	Secondary: West
Panel A. Subsample estimates						
Lagged HSR access	0.109	0.454	1.487**	-0.379	-0.600	-2.145***
	(0.477)	(0.485)	(0.713)	(0.512)	(0.522)	(0.794)
Within R^2	0.775	0.803	0.757	0.687	0.676	0.634
N	933	949	682	933	949	682
Panel B. Direct regional effects						
HSR access × East	-0.144	-	-	-0.030	-	-
	(0.501)	-	-	(0.534)	-	-
HSR access × Central	-	1.621***	-	-	-1.142**	-
	-	(0.458)	-	-	(0.482)	-
HSR access × West	-	-	1.388**	-	-	-2.550***
	-	-	(0.609)	-	-	(0.676)
Central = East (*p*-value)	-	0.007	-	-	0.102	-
West = East (p-value)	-	-	0.044	-	-	0.002
West = Central (*p*-value)	-	-	0.745	-	-	0.076
N (pooled sample)	2,564	2,564	2,564	2,564	2,564	2,564

Note: Panel A estimates separate regional models, allowing control coefficients and year effects to vary. Panel B jointly estimates Equation (2); the reported p-values test equality across regions. Because the panels impose different coefficient restrictions, their point estimates need not coincide. All models include the full controls, city FE, and year FE; city-clustered standard errors are in parentheses. FE = Fixed Effects. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels respectively.

4.2. Extensive and intensive margins

The distinction between first access and subsequent connectivity is central to the paper's policy question. This subsection therefore compares the extensive and intensive margins on a common estimation sample.

Table 4 reports four specifications on a common sample of 2,564 observations: (1) L.addlines alone; (2) L.addlines and L.hub; (3) L.hsrdummy and L.addlines; and (4) all three variables. L.addlines counts lines beyond the first and is missing whenever the lagged line count or HSR indicator is missing.

In column (4), first HSR access raises the tertiary share by 0.932 points and reduces the secondary share by 1.081 points (both $p < 0.01$). By contrast, additional-line effects are imprecise (tertiary -0.100, $p = 0.798$; secondary -0.025, $p = 0.951$), as are hub-status effects (tertiary -0.120, $p = 0.881$; secondary -0.868, $p = 0.323$). Formal tests comparing first-access and additional-line coefficients yield $p = 0.054$ for tertiary and p = 0.073 for secondary. Large, precise first-access effects and near-zero line-addition estimates point to the extensive margin as economically dominant; the direct contrasts are statistically borderline at conventional cutoffs.

Economically, this pattern supports the interpretation that entering the national passenger network is more consequential than adding further connections after a city has already joined the network. The available data do not directly observe passenger flows, commuting, or firm relocation, so the exact channels remain unclear. The imprecise line-count and hub estimates should not be read as zero effects and do not preclude effects on other dependent variables, such as housing prices, wages, or innovation.

Table 4. Decomposing HSR expansion: extensive and intensive margins

Variable	Tertiary (1)	Tertiary (2)	Tertiary (3)	Tertiary (4)	Secondary (1)	Secondary (2)	Secondary (3)	Secondary (4)
L.hsrdummy	-	-	0.935***	0.932***	-	-	-1.060***	-1.081***
			(0.327)	(0.328)			(0.353)	(0.355)
L.addlines	-0.094	-0.012	-0.130	-0.100	-0.290	-0.127	-0.249	-0.025
	(0.268)	(0.385)	(0.267)	(0.389)	(0.266)	(0.407)	(0.266)	(0.417)
L.hub	-	-0.321	-	-0.120	-	-0.635	-	-0.868
		(0.789)		(0.798)		(0.865)		(0.878)
Within R^2	0.755	0.755	0.757	0.757	0.631	0.631	0.634	0.634
p-value: access = addlines	-	-	0.015	0.054	-	-	0.081	0.073
N	2,564	2,564	2,564	2,564	2,564	2,564	2,564	2,564

Note: All models use the same 2,564 observations and include the full controls, city FE, and year FE. City-clustered standard errors are in parentheses. L.addlines > 0 in 329 observations; L.hub = 1 in 107. ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels respectively.

4.3. Primary-sector implications

The primary-sector share provides a complementary compositional comparison. Because the three sectoral GDP shares are mechanically linked, however, this regression is neither an independent placebo nor a direct test of labor mobility, land acquisition, or urbanization mechanisms.

In the full specification (N = 2,564), the primary-sector coefficient is 0.151 (p = 0.372). The central estimate, 0.144 (p = 0.611), and eastern estimate, 0.269 (p = 0.191), are not statistically significant. The western estimate is positive and statistically significant at the 5 percent level (0.654, p = 0.047), but its magnitude is smaller than the western tertiary estimate (1.487) and the absolute western secondary estimate (2.145). These comparisons are descriptive rather than formal tests of sectoral differences.

The cohort comparison also yields a positive primary-sector coefficient of 0.757 (p = 0.050, rounded). Because the sectoral shares sum to approximately 100 percent, these results are best read as a compositional complement to the tertiary- and secondary-sector estimates rather than as an independent mechanism test.

5. Robustness and endogeneity

5.1. Robustness to sample composition and alternative fixed effects

Because provincial capitals, directly administered municipalities, and cities with independent planning status differ systematically in administrative resources and HSR placement, the baseline is reestimated after excluding them. The effects remain statistically significant (tertiary 0.845, $p < 0.05$; secondary -0.940, $p < 0.05$). Because fiscal-expenditure missingness is concentrated in 2011, the full specification is also reestimated without that year. The resulting sample contains 2,526 observations, with estimated effects of 0.911 for tertiary and -1.086 for secondary (both $p < 0.01$); their directions and magnitudes are similar to the baseline estimates.

Replacing year fixed effects with province-by-year fixed effects absorbs province-level time-varying shocks. The secondary-sector effect remains significant (-0.697, $p < 0.05$), whereas the tertiary estimate falls to +0.450 and becomes imprecise (p = 0.126); the primary-sector estimate is +0.244 (p = 0.174). Adding city-specific linear trends produces the same ranking: secondary -0.666 ($p < 0.05$) and tertiary +0.419 (p = 0.204). Accordingly, secondary-sector contraction is the paper's most robust structural effect, while tertiary expansion is more sensitive to differential-trend controls.

5.2. Pretreatment trends and placebo tests

A valid timing placebo must exclude all observations after actual HSR opening. The placebo therefore retains only untreated city-years and never-treated cities and assigns pseudo-treatment to the three years immediately preceding first HSR access. The resulting sample contains 1,586 observations from 259 cities. Estimated pseudo-treatment coefficients are 0.335 (p = 0.333) for tertiary, -0.323 (p = 0.399) for secondary, and -0.010 (p = 0.965) for primary. The absence of detectable pseudo-treatment effects strengthens the timing-based causal interpretation of the main results, while not eliminating every form of anticipation or time-varying confounding.

5.3. 2014-cohort difference-in-differences comparison

A complementary cohort-based Difference-in-Differences (DID) design compares cities first connected in 2014 with cities never connected during the observed period. Before complete-case restrictions, the comparison includes 29 treated cities and 90 never-treated cities. The estimation sample contains 27 treated cities, 81 never-treated cities, and 1,008 observations after requiring nonmissing outcomes and controls. Post-treatment begins in 2015, consistent with the one-year lag. The specification includes the full control set, city and year fixed effects, and city-clustered standard errors; the event study uses 2013 as the omitted year. Joint tests show no detectable differential pretreatment trends (tertiary $F = 1.10$, $p = 0.366$; secondary $F = 0.68$, $p = 0.640$), supporting identification in this cohort. Because the full-sample Two-Way Fixed Effects (TWFE) model pools cities connected at different dates, it may combine heterogeneous cohort responses. The cohort design therefore provides a focused causal comparison rather than a complete correction for staggered adoption. Figure 1 presents the secondary-sector event-study estimates.

The cohort DID estimates show that 2014 HSR access raises the tertiary share by 2.148 points and reduces the secondary share by 2.907 points (both $p < 0.01$). Western effects are +3.278 and -4.493 points (both $p < 0.01$). These larger effects are compatible with the full-sample results because responses can differ by cohort, region, and exposure duration [13]. The treated sample includes 13 central, 10 western, and 4 eastern cities, so cohort composition matters for the magnitudes. Overall, the comparison strengthens the evidence that first access shifts sectoral composition away from secondary and toward tertiary activity, although the estimates apply to a distinct treatment group.

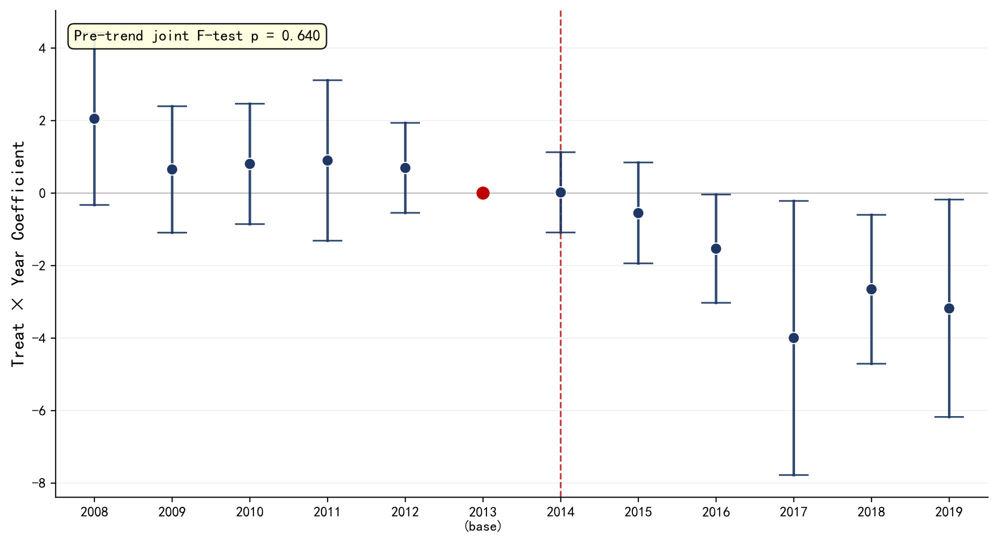

Figure 1. Event study of the secondary-sector GDP share (2014 Cohort)

Note: The vertical dashed line marks the HSR opening year, 2014; the lagged post-treatment period begins in 2015. The omitted year is 2013. Error bars show 95% confidence intervals. The jointly insignificant pretreatment coefficients (tertiary $p = 0.366$; secondary $p = 0.640$) support the parallel-trends assumption for this cohort comparison, subject to sampling uncertainty.

5.4. Endogeneity concerns and sensitivity

Identification relies on within-city treatment timing rather than random route placement. Economic potential, administrative status, and geography may influence route selection, creating risks from reverse causality and time-varying omitted variables. City fixed effects remove time-invariant selection, and province-by-year effects absorb shared provincial policy shocks. The pretreatment-only placebo, event study, and cohort DID then test whether the observed timing patterns are compatible with a causal interpretation. None of these checks makes treatment assignment random, but together they provide a structured assessment of pretrends and selection. The same-sample comparisons below isolate how the target coefficient responds to observed controls and complement these design-based checks.

For the national complete-case sample ($N = 2,564$), the tertiary coefficient changes from +0.897 without controls to +0.925 with controls, while the secondary coefficient changes from -1.100 to -1.078. The primary coefficient changes from +0.201 to +0.151, and its full-model estimate is not significant ($p = 0.372$). Holding the sample fixed removes movement caused by observations entering or leaving the regression. The modest tertiary and secondary changes show that the effects are stable to included observables rather than driven by sample composition. This strengthens confidence in the baseline estimates, although time-varying unobservables remain a source of uncertainty.

For western cities, both restricted and full models use the same 682 observations. The tertiary coefficient changes from 1.067 to 1.487, and the secondary coefficient changes from -1.245 to -2.145. The signs are unchanged, while the larger coefficient movements indicate that the exact western magnitudes are more sensitive to the included controls than the national estimates.

Across sample restrictions, placebo timing, alternative fixed effects, cohort comparisons, and coefficient-stability checks, the evidence supports a robust negative effect of first HSR access on the secondary-sector share. The positive tertiary effect is clear in the baseline and cohort DID but loses precision under demanding trend controls; primary-sector results vary. The central causal result is therefore structural reallocation whose most robust component is secondary-sector contraction, with tertiary expansion as complementary evidence.

6. Conclusion

This paper estimates how HSR expansion affects urban industrial composition through first network access and subsequent line and hub expansion. Using an unbalanced panel of Chinese prefecture-level cities from 2008 to 2019, the complete-case two-way fixed effects sample contains 282 cities and 2,564 observations. First HSR access raises the tertiary GDP share by 0.925 percentage points and reduces the secondary share by 1.078 percentage points. Pooled interaction estimates show larger tertiary effects in central and western cities than in the east, while the secondary-sector effect is most pronounced in the west. After initial access is controlled for, additional lines and hub status have no precisely estimated average effects, and formal comparisons between the two margins are statistically borderline. The secondary-sector contraction remains significant under province-by-year effects and city-specific trends; the tertiary expansion is more sensitive to those demanding specifications. A pretreatment-only placebo and the 2014-cohort comparison support the timing and direction of the main effects. Taken together, the evidence identifies first network access as the principal margin through which HSR expansion reshapes urban industrial composition. The larger first-access estimates imply that connecting previously unserved cities generates stronger structural adjustment, especially outside the east. This finding does not, by itself, establish a cost-benefit ranking: investment decisions should also weigh construction and operating costs, utilization, connectivity quality, welfare, distributional effects, and outcomes beyond sectoral GDP shares.

Several limitations remain: route placement is nonrandom; some contemporaneous controls may respond to prior HSR access; segmented openings may introduce measurement error; and sectoral GDP shares do not reveal specific mechanisms. These issues qualify the exact magnitudes and mechanism interpretation. Future work could combine staggered-adoption estimators with employment, passenger-flow, or firm-level data to test mechanisms and broader welfare effects.

References

[1] Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *The Quarterly Journal of Economics, 131*(2), 799–858. https://doi.org/10.1093/qje/qjw002

[2] Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies, 81*(3), 1046–1070. https://doi.org/10.1093/restud/rdu010

[3] Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics, 145*, Article 102442. https://doi.org/10.1016/j.jdeveco.2020.102442

[4] Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *The Review of Economics and Statistics, 99*(3), 435–448. https://doi.org/10.1162/REST_a_00660

[5] Sun, W., Niu, D., & Wan, G. (2022). Transportation infrastructure and industrial structure upgrading: Evidence from China's high-speed railway [In Chinese]. *Management World, 38*(3), 19–34. https://doi.org/10.19744/j.cnki.11-1235/f.2022.0045

[6] Deng, H., Yang, L., & Pan, X. (2020). Can high-speed rail help upgrade industrial structure? Facts and mechanisms [In Chinese]. *Journal of Finance and Economics, 46*(6), 34–48. https://doi.org/10.16538/j.cnki.jfe.2020.06.003

[7] Bian, Y., Wu, L., & Bai, J. (2019). Does high-speed rail improve regional innovation in China? [In Chinese]. *Journal of Financial Research*, (6), 132–149.

[8] Qiao, B., Zhang, R., & Lei, C. (2019). HSR effects, producer services agglomeration, and manufacturing upgrading [In Chinese]. *Economic Review*, (6), 80–96. https://doi.org/10.19361/j.er.2019.06.06

[9] Department of Urban Socioeconomic Survey, National Bureau of Statistics of China. (2009–2020). *China city statistical yearbook* [Annual series]. China Statistics Press.

[10] National Development and Reform Commission, Ministry of Transport, & China Railway Corporation. (2016, July 13). *Medium- and long-term railway network plan* [Policy plan; in Chinese]. https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/201607/t20160720_962188.html

[11] Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

[12] Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy, 99*(3), 483–499. https://doi.org/10.1086/261763

[13] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001