ISSN 2616-5902 Vol. 6, Issue 9: 61-68, DOI: 10.25236/AJBM.2024.060909

The Impacts of High-Speed Railways on Cities with Difficult Terrains in China

Auston Jiaqiang Yang[1,a,*], **Carson Jiasheng Yang**[1,b]

[1] *American International School of Guangzhou, Guangzhou, China*
[a] *austonjqyang@hotmail.com*, [b] *carsonjsyang@hotmail.com*
[*] *Corresponding author*

Abstract: *Significant disparities among regions characterize economic development in China. While some cities have experienced rapid growth and improved living standards, others lag. This paper investigates whether the expansion of High-Speed Rail (HSR) and increased connectivity with wealthier regions have facilitated catch-up growth, particularly in disadvantaged areas by geography. To assess the causal impact, we exploit the staggered opening of HSR stations and employ stacked Differences-in-Differences as our identification strategy. Recognizing that the expansion of HSR stations is not random, we use propensity score matching (PSM) to identify similar control regions. We categorize Chinese counties into three samples for PSM: the full sample, less developed counties, and counties with challenging terrain. Our findings suggest that an HSR station significantly boosts economic activity, as measured by nighttime lights (NL), even in regions with difficult terrain. We also use urban cover, derived from satellite imagery, as a proxy for urban development. Our results indicate that HSR openings lead to a significant increase in urban expansion in less developed regions but do not significantly impact urbanization in regions with challenging terrain.*

Keywords: *economic development, transportation, urban economics*

1. Introduction

Economic development tends to be spatially uneven within countries, with some regions experiencing rapid growth while others lag. The reasons for this disparity can vary, including lack of policy attention, an educated labor force, and business opportunities, all of which contribute to some cities' inability to grow quickly or catch up due to a lack of productive industries. Even more puzzling is that some regions, despite being close to wealthier areas, still struggle to progress. Proximity alone does not guarantee access to knowledge, labor, or business opportunities, especially in China, where many cities remain disconnected due to remoteness and infrastructure constraints, including challenging terrains that hinder the construction of roads or railways. This lack of connectivity obstructs the flow of knowledge, labor, and economic opportunities. This paper examines whether infrastructure improvements and increased connectivity with the rest of the country can stimulate economic growth in less-developed regions, particularly those with challenging terrains.

We analyze the expansion of High-Speed Railways (HSR) and its impact on less developed regions and areas with difficult terrains. HSR expansion significantly improves commuting and goods transport, with trains traveling at speeds exceeding 250km/h. Between 2008 and 2022, China rapidly expanded its HSR network, constructing over 1424 stations in 802 counties. Using data on HSR openings, economic and geographic characteristics, and economic indicators at the county level from 2000 to 2019, we assess the impact of HSR expansion on these regions. Given that HSR expansion is not random, we use propensity score matching (PSM) to identify comparable control counties for those with early access to HSR stations. Additionally, to account for time-variation impacts, we employ stacked differences-in-differences (DiD) for causal identification. Our analysis relies on the granularity and high frequency of nighttime lights (NL) and land cover classifications to measure economic outcomes. Our findings indicate that HSR expansion has a significant positive impact on both less developed counties and those with challenging terrains. It increases NL by 13.73% for less developed counties and 8.61% for counties with difficult terrains. HSR expansion also promotes urbanization in less developed counties, though it does not have the same effect on counties with challenging terrains.

In this study, we compile a comprehensive dataset tailored to address our research questions, focusing specifically on less developed regions and regions with challenging terrain. To define difficult terrain, we calculate three variables within each county: median elevation, average slope, and the share of land with a slope exceeding 15 degrees. We categorize counties with a median elevation exceeding 126.26 meters as having difficult terrain. Less developed regions are identified as those with nighttime lights (NL) values lower than the 75th percentile cutoff for the years 1995-1999. To ensure comprehensive matching, we collect two types of variables. In addition to elevation and slope, we consider crop and forest cover to capture the natural endowment of a location. Using census data, we also assess human capital and economic conditions in the year 2000, including variables such as population density, the share of the population with a college education, and the share of labor in manufacturing. Finally, we incorporate two economic outcome variables spanning two decades: NL to proxy economic activities and urban cover to measure the degree of urbanization. All variables are harmonized and aggregated at the county level to examine the impact of High-Speed Rail (HSR) expansion on county-level outcomes.

To assess the impact of High-Speed Rail (HSR) expansion, we employ a combination of Propensity Score Matching (PSM) and stacked Differences-in-Differences (DiD) as our identification method. The decision for a county to receive an HSR station early is likely not random, influenced by various factors. Therefore, we meticulously control for a comprehensive list of variables to construct an appropriate control group for matching. The key identification assumption is that counties with early HSR station openings would have followed parallel trends as their matched counterparts without HSR expansion. Event study plots later provide empirical support for this assumption. We conduct PSM on three samples to address our research questions: the full sample, less developed counties, and high-elevation counties. Subsequently, we employ the stacked DiD strategy instead of the conventional two-way fixed effects model. This decision is based on the likelihood of time-variation impacts from HSR expansion. The effects of HSR introduction could be significant initially and then taper off, or they may persist over time.

First, we analyze the results of Propensity Score Matching (PSM). Our findings indicate that flatter counties with higher population densities, more educated workforces, and greater manufacturing activities are more likely to have an HSR station. Following PSM, we assess the density plots and balance table for counties with early HSR station openings (``treated") versus those with late openings (“control”). Before matching, the distribution of control counties differs significantly from treated counties. Treated counties tend to have higher average NL radiance in 1995-1999, denser populations, greater crop coverage, lower elevation, higher education levels, more workers in manufacturing, and flatter terrains. However, post-PSM, the density plots for control counties closely resemble those for treated counties. This matching is particularly effective for the less developed and high-elevation samples. Following PSM in the full sample, the differences between treated and control counties diminish. For less developed regions after matching, there is virtually no difference in any characteristics between treated and control counties. Similarly, in the high-elevation sample, matched treated and control counties exhibit striking similarities.

Having established appropriate control counties, we proceed to examine event study plots. These plots indicate that, for all three samples, there are no discernible pre-trends, supporting the parallel trends assumption. They also demonstrate that treated counties experience significant and sustained increases in NL from the year of station opening. Regression analysis reveals that early HSR station openings lead to a 13.73% increase in NL for less developed counties and an 8.61% increase for counties with challenging terrains. However, the impact of HSR expansion on urbanization is less pronounced. Early access to an HSR station is associated with a modest 1.87% increase in urban cover for less developed counties, which is only weakly significant. In contrast, for high-elevation counties, we find that HSR stations have no significant impact on the conversion of land to built-up areas. These results suggest that, despite major infrastructure improvements such as HSR, geographical disadvantages faced by high-elevation counties remain significant barriers to urbanization. Additional efforts may be needed to address these challenges.

The paper mainly has three contributions. Firstly, it adds to the literature on the relationship between transportation access and regional growth. While previous studies have addressed similar questions about railroads[1] or highways[2], this paper contributes with novel data and by examining a significant shift in transportation modes. Secondly, it contributes to the ongoing literature on the impacts of High-Speed Rail (HSR). While many papers have studied its impacts on economic activities[3], urban access[4], urban specialization[5], regional equity[6], pollution[7] among others, this paper introduces a new research design. Propensity Score Matching (PSM) is applied to create a more comparable sample, and Stacked Differences-in-Differences (DiD) is used to address concerns about estimating time-variant treatment effects. Lastly, the paper highlights significant findings regarding the challenges faced by regions with disadvantaged geography in their development. While some papers have discussed the costs of difficult geography from a theoretical standpoint[8] or stagnant economic growth faced by landlocked countries[9], this paper examines a large infrastructure project where the Chinese government invested substantial financial and human resources in building high-speed railways in remote areas with challenging terrain. The findings suggest that despite increased infrastructure and connectivity, these regions are developing at a slower pace.

2. Data & Empirical Strategy

To investigate the research question in this paper, we construct a dataset encompassing 2,408 counties in China from the years 2000 to 2019. The opening dates of HSR stations are manually collected from the official railway website (www.12306.cn) and China's Ministry of Railways. In addition to their opening year, we gather information on each HSR station's name, prefecture, and province. Using OpenStreetMap, we geocode the exact latitude and longitude of stations, enabling us to map each station to its corresponding county. We obtain the 2000 census data and county (ADM3-level) shapefile from IPUMS International. This dataset provides information on the industries in which the labor force was employed, educational attainment, and population density at the county level in 2000. These variables are instrumental in the Propensity Score Matching (PSM) process to identify comparable control counties for those with early HSR station openings. The shapefile contains boundaries for 2,406 counties, which are used to calculate average nighttime lights, the proportion of land covered by forest, crops, or urban areas, average terrain slope, elevation, and the proportion of steep terrain.

Nighttime lights data from 2000 to 2020 are sourced from the harmonized dataset[10]. The harmonized NL data are used for two main reasons. First, our study period spans from 2000 to 2019, during which two satellites, DMSP or VIIRS, were used to measure nighttime light radiance. These satellites cover different periods and employ varying resolutions and measurement methods. By using harmonized NL data, we can incorporate both sources of data simultaneously. Second, even for the years covered by the DMSP satellite, temporal adjustments are necessary. Therefore, we opt to use the harmonized NL dataset for consistency and accuracy across the study period. Urban, forest, and crop covers are determined using land cover classifications. The data utilized is the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data product (Friedl et al., 2019). This dataset provides global land cover types annually from 2001 to 2020 at a resolution of 500 meters. The classification algorithm is supervised, and additional post-processing is conducted to further refine specific classes.

2.1. Visualizing NL & HSR Data

Figure 1 and 2 presents maps illustrating county-level NL for the years 2000 and 2019. The maps show an increase in overall brightness over time, indicating increased economic activity. In 2000, economic activities were primarily concentrated around coastal cities, while by 2019, they had expanded westward and inland.

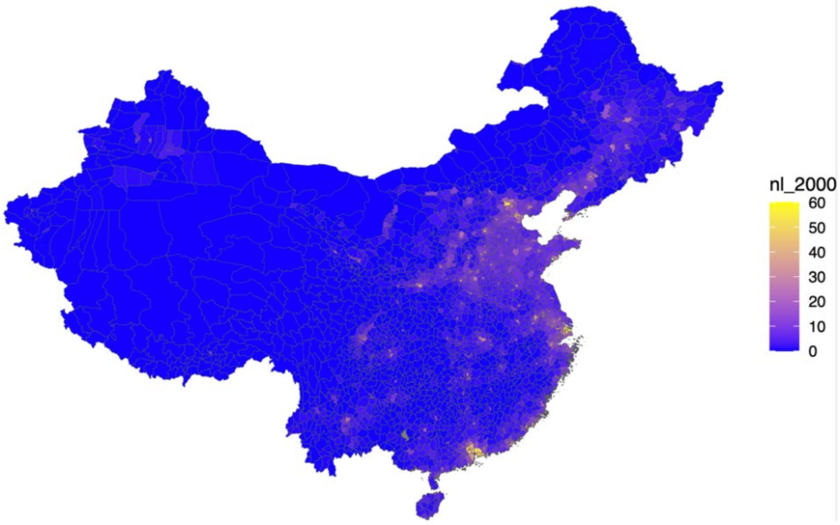

Figure 1: NL in China in 2000.

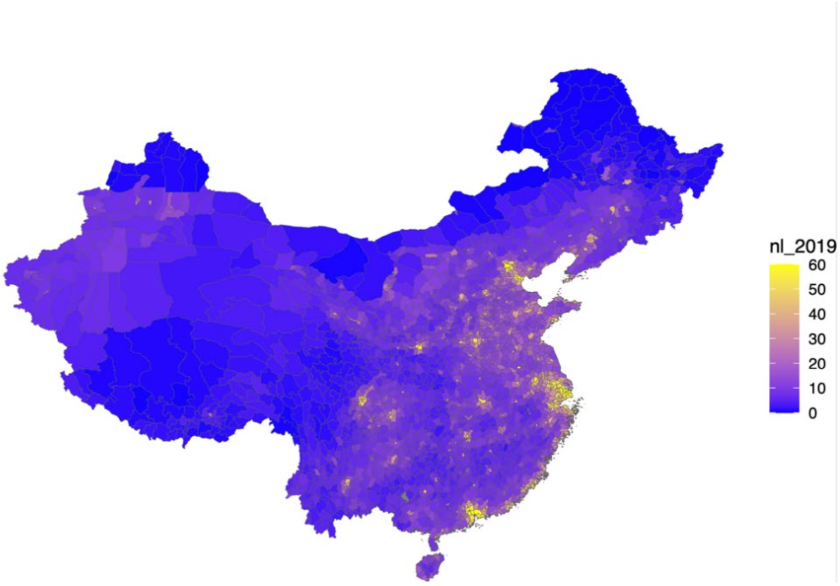

Figure 2: NL in China in 2019.

Figure 3 displays the locations of HSR stations. Blue diamonds mark HSR stations that opened on or before 2018. Green triangles mark HSR stations that opened after 2018. Stations that opened early primarily connect northern and southern regions, as well as coastal areas. In contrast, those that opened later tend to connect more inland locations.

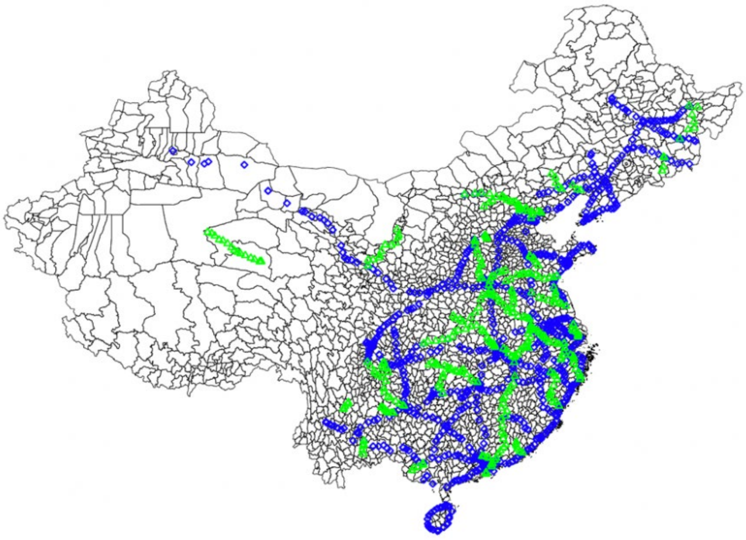

Figure 3: HSR stations.

2.2. Empirical Strategy

To establish the causal impacts of HSR station openings on local economic development, we utilize a combination of Propensity Score Matching (PSM) and Stacked Difference-in-Differences (DiD). There are two primary concerns for ensuring causal interpretations. Firstly, the selection of counties to have an HSR station is typically non-random and may be influenced by economic conditions, population characteristics, and geographic proximity to major cities, among other factors. Given these differences, it is unlikely that treated and control regions would have exhibited parallel trends before HSR opening. Simply comparing counties with HSR stations to those without would result in biased causal estimates. To address this, we use PSM to identify similar control counties that would have experienced parallel changes over time in the absence of HSR stations. Secondly, since HSR openings are staggered, there may be time-varying treatment effects. To account for this, we employ stacked DiD. The key assumption here is that counties receiving HSR stations earlier and those receiving them later would have experienced parallel changes over time if the treatment (i.e., HSR opening) had not occurred.

2.2.1. Propensity Score Matching

For Propensity Score Matching (PSM), treatment is defined as having an HSR station open on or before the year 2018. This threshold is chosen because by 2018, 25% of Chinese counties had been connected to HSR. Therefore, counties connected to HSR after 2018 are considered controls. The selection of counties to have an HSR station open early is likely non-random. These counties may serve as regional hubs or be strategically connected to link two major cities. The rationale behind PSM is to identify control counties that are similar to treated counties but have a later opening year not due to factors related to economic outcomes. There is a concern that more developed regions are getting connected to HSR early and may have experienced greater growth even without HSR opening. To address this, we limit our sample to less developed regions. Less developed regions are defined as counties with nighttime lights (NL) values lower than 9 during 1995-1999, which corresponds to the 75th percentile NL for treated regions. Additionally, counties with difficult terrains, such as mountainous areas, tend to have later HSR openings due to engineering challenges. These regions are also typically less developed, as their challenging terrain makes road network construction similarly difficult. Thus, we conduct PSM on counties with elevations over 139.5 meters, which is the median elevation among all treated counties. PSM is performed using nearest neighbor matching and logit regressions to calculate distances. Treated counties are matched with similar control counties, and if no match is found, both treated and control counties are discarded. Additionally, the same control county may be matched with multiple treated counties.

2.2.2. Stacked Difference-in-Differences

Given the staggered nature of HSR openings, this paper employs the stacked Difference-in-Differences (DiD) approach as its identification strategy. This method has been utilized in several published papers[11] as the identification strategy. The key assumption is that counties receiving HSR stations early would have exhibited parallel trends to those receiving stations later if the treatment had not been implemented.

3. Results

3.1. Propensity Score Matching

The timing of HSR station openings is likely not random, as geographic constraints can immediately pose challenges to railway construction. Mountainous regions, for instance, may require the construction of tunnels or large bridges to connect railway stations. Moreover, the selection of railway station locations involves a process influenced by supply and demand factors. Areas with denser populations and greater business activities naturally exhibit a higher demand for high-speed trains compared to regions with lower population densities. From a government planning perspective, policymakers may prioritize connecting major cities first to optimize economic growth.

The results in Table 1 confirm these hypotheses. Columns (1), (2), and (3) demonstrate robust findings indicating that counties with a higher proportion of steep terrain and a higher share of land dedicated to crops are more likely to receive an HSR station later. Conversely, counties with denser populations, a higher proportion of college graduates, and more manufacturing activities are more likely to receive an HSR station early. This pattern persists even after including province and prefecture-fixed effects. In China, there are significant variations in terrain and development levels among provinces. To ensure that the regressions do not merely reflect the fact that some provinces receive HSR stations earlier than others, it is important to examine within-province or even within-prefecture variations. Our robust results in column (4) indicate that within a province, flatter counties with denser populations, more educated workforces, and more manufacturing activities are more likely to receive an HSR station.

Using the same set of characteristics, we applied PSM to identify comparable control counties. We performed PSM on three samples: the full sample, less developed counties, and counties with difficult terrains. Figure 4 presents density plots before and after PSM for counties that had an HSR station open earlier vs. later. Before any matching was done, the distribution of control counties differed significantly from that of treated counties in the full sample, indicating differing characteristics. However, after PSM, the density plots for control counties became quite similar to those for treated counties. Visually, the matching appears to be better for the less developed and high-elevation samples. The density plots for treated and control counties exhibit similar long right tails, suggesting that matched control counties are comparable to treated ones.

Table 1: Factors that influenced HSR openings.

	(1)	(2)	(3)	(4)	(5)
Share Steep	1.04***	2.18***	2.01***	1.70*	1.73*
	(0.379)	(0.647)	(0.645)	(0.933)	(0.954)
Elevation(km)	1.05***	0.095	0.115	0.024	0.716*
	(0.097)	(0.161)	(0.166)	(0.25)	(0.427)
Forest Cover (2001)		-0.005	-0.004	0.006	0.006
		(0.005)	(0.005)	(0.006)	(0.008)
Crops Cover (2001)		0.021***	0.017***	0.01	0.007
		(0.004)	(0.004)	(0.006)	(0.007)
Ln(PopDen2000)		-0.862***	-0.766***	-0.597***	-0.878***
		(0.108	-0.109)	(0.185)	(0.181)
NL (1995-1999)		-0.069***	0.002	-0.037	-0.002
		(0.018)	(0.02)	(0.035)	(0.035)
% College			-30.1***	-16.3**	-24.5***
			(6.13)	(6.12)	(7)
% Above Highsch			9.51***	1.49	4.69
			(2.85)	(3.53)	(3.88)
% workers in Agri			-1.43	-2.75*	-3.78***
			(1.24)	(1.53)	(1.45)
% workers in Manu			-11.9***	-11.1***	-12.1***
			(1.98)	(2.59)	(2.87)
Observations	2,408	2,408	2,380	2,380	2,380
R2	0.07829	0.16019	0.19365	0.25129	0.39184
Province FE	N	N	N	Y	N
Prefecture FE	N	N	N	N	Y
Note: *** p$<$0.01, ** p$<$0.05, * p$<$0.1					

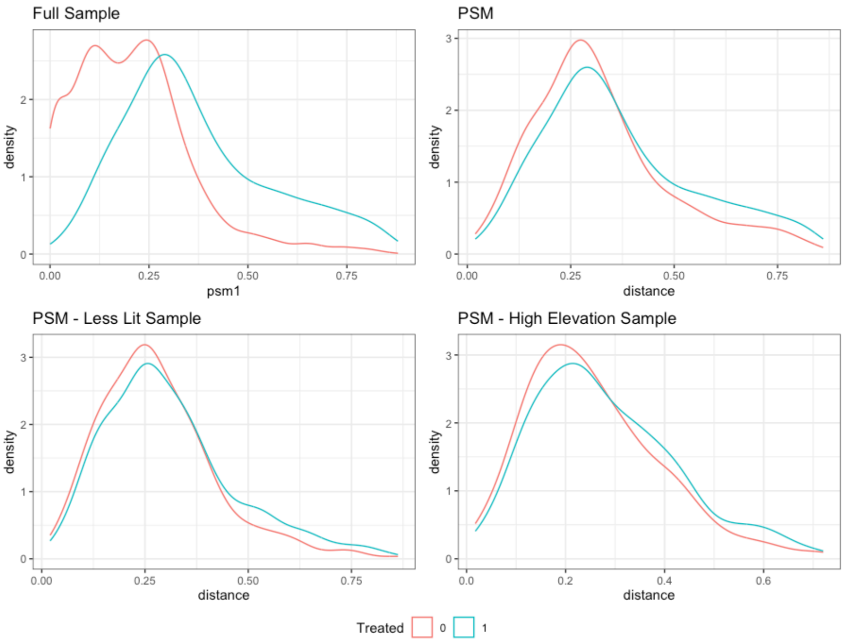
Full Sample
PSM
PSM - Less Lit Sample
PSM - High Elevation Sample

Figure 3: HSR stations.

3.2. Effects of HSR opening on economic activities

Table 2 shows the magnitude of the impacts of HSR openings on economic activities. Column (1) in Table 2 indicates that early HSR station openings have a significant and positive impact on economic activities. Considering the average NL radiance is 10.31, having an HSR station increases NL by 4.07%. In column (2), we observe the effects of HSR openings on less developed counties, indicating large, statistically significant, and positive impacts from HSR openings. For less developed counties, those with early HSR stations experience a 13.73% increase in NL. Compared with the PSM full sample, the magnitude of impacts on less developed regions is much larger. These findings suggest that in less developed regions, the impacts of HSR on economic activities could be larger compared to more developed regions. Findings for high-elevation counties are similar. The impacts of HSR openings are positive on economic activities, increasing NL by 8.61%. The magnitude of impacts, although larger than the PSM sample, is smaller than the less developed counties sample. Combining all the results, we find that the impacts of opening an HSR station are generally positive on economic activities, especially in less developed regions. From the descriptive statistics, we already know that high-elevation counties are of similar development level as the less developed counties, perhaps even less developed than them. However, the impacts of HSR opening are not as large for high-elevation counties, suggesting that geographical disadvantages could dampen the positive impacts of a large infrastructure push.

Table 2: Impacts of HSR openings on NL.

	PSM	PSM-Less Develop	PSM-High Elevation
	(1)	(2)	(3)
Post x Treated	0.407*	0.759**	0.476*
	(0.205)	(0.297)	(0.268)
Observation	134842	110566	74566
Adjusted R-squared	0.97	0.914	0.957
Incl Controls	Y	Y	Y
Dep Var Mean	10.31	5.525	5.524
Note: *p < 0.1; **p < 0.05; ***p < 0.01			

3.3. Effects of HSR opening on urban settlement

Nighttime lights only measure one type of economic outcome. To provide a more comprehensive picture of HSR's impacts on regional economic growth, we use log urban cover to measure urbanization. We use the log transformation instead of the level of urban cover because the distribution of urban cover has a long right tail, i.e., many counties with 100\% urban land. Zeros are adjusted to 0.01 in the log transformation. Urban cover indicates the extent of land that has been developed into human settlements, and its detection is done using algorithms that analyze satellite imagery. This method also allows us to determine the extent of land covered by water bodies, grass, forests, crops, and so on. Remote sensing data and methods work well because built-up areas look very different both visually and numerically, allowing us to observe urbanization progression from space. Leveraging the granularity of this data, we examine the impacts of HSR openings on urbanization.

Table 3: Impacts of HSR openings on Urban Cover.

	PSM	PSM-Less Develop	PSM-High Elevation
	(1)	(2)	(3)
Post x Treated	0.0218**	0.0187	0.00993
	(0.0102)	(0.0119)	(0.0144)
Observation	126693	103927	70084
Adjusted R-squared	0.998	0.997	0.997
Incl Controls	Y	Y	Y
Dep Var Mean	0.888	0.343	0.0873
Note: *p < 0.1; **p < 0.05; ***p < 0.01			

Column (1) in Table 3 demonstrates that receiving an HSR station early significantly increases urbanization by 2.18%. Though HSR opening has positive impacts on urbanization, the magnitude of impacts is smaller compared to its impacts on NL. For less developed counties, receiving an HSR station early increases urban cover by 1.87%. The increase is weakly significant. Lastly, for high-elevation counties, we discover that HSR stations have no impact on whether more land is being converted to built land. Compared to the impacts of HSR opening on NL, its impacts on urbanization are more muted. We still observe that getting HSR stations early allows less developed regions to become more urbanized. The results are interesting in that high-elevation counties are not affected by HSR. It is worth noting that less developed and high-elevation counties are way less urbanized to begin with. On average, the PSM sample has 10% urban cover, less developed counties have 3.4% urban cover, and high-elevation counties have 4.4%. These results suggest that even major infrastructure improvements like HSR would not be sufficient to overcome the geographical disadvantages that high-elevation counties face. They still struggle with challenges in becoming more urbanized.

4. Conclusions

This paper examines the effects of High-Speed Rail (HSR) openings on regional economic outcomes. Our methodology combines a stacked Difference-in-Differences (DiD) strategy to estimate time-variant effects and Propensity Score Matching (PSM) to create comparable treatment and control groups. We focus not only on all counties in China but also specifically on those that are less developed and have difficult terrains. Our analysis uses nighttime lights and urban cover as measures of regional economic outcomes, providing granular, yearly data. Matching is conducted on a comprehensive set of characteristics to address concerns regarding the non-random timing of HSR openings. This approach ensures that treated counties closely resemble control counties, minimizing endogeneity concerns.

The construction of High-Speed Rail (HSR) in China is a significant infrastructure development aimed at reducing travel times and promoting economic exchanges between regions. Our analysis shows that HSR openings have had a notably positive impact on economic activities, as indicated by a 13.73\% increase in nighttime lights in less developed regions and an 8.61\% increase in high-elevation areas. However, the impact is less pronounced in high-elevation counties. Despite receiving the same infrastructure improvement, these counties are not developing as rapidly as similar, but lower-elevation regions. This disparity is even more evident in urbanization rates, where high-elevation counties with HSR stations are not more urbanized than those without. These findings underscore the notion that while HSR openings offer some benefits, they are not sufficient to overcome the developmental challenges faced by regions with difficult terrains.

Further research could be conducted to examine the economic development of disadvantaged regions further. This paper focuses on a limited set of outcomes, but other factors such as education attainment and healthcare also matter. Elevation and the share of steep land are not the only indicators of regions facing disadvantages. Lastly, urbanization might be less critical for regions with difficult terrain, as building there may not be cost-effective. More work are required to determine the optimal growth path for regions with challenging terrains.

References

[1] Baum-Snow N, Brandt L, Henderson JV, Turner MA, Zhang Q. Roads, railroads, and decentralization of Chinese cities. Review of Economics and Statistics. 2017 Jul 1; 99(3):435-48.
[2] Banerjee A, Duflo E, Qian N. On the road: Access to transportation infrastructure and economic growth in China. Journal of Development Economics. 2020 Jun 1; 145:102442.
[3] Diao M. Does growth follow the rail? The potential impact of high-speed rail on the economic geography of China. Transportation Research Part A: Policy and Practice. 2018 Jul 1; 113:279-90.
[4] Ke X, Chen H, Hong Y, Hsiao C. Do China's high-speed-rail projects promote local economy?—New evidence from a panel data approach. China Economic Review. 2017 Jul 1; 44:203-26.
[5] Lin Y. Travel costs and urban specialization patterns: Evidence from China's high speed railway system. Journal of Urban Economics. 2017 Mar 1; 98:98-123.
[6] Zhang F, Yang Z, Jiao J, Liu W, Wu W. The effects of high-speed rail development on regional equity in China. Transportation Research Part A: Policy and Practice. 2020 Nov 1; 141:180-202.
[7] Lin Y, Qin Y, Wu J, Xu M. Impact of high-speed rail on road traffic and greenhouse gas emissions. Nature Climate Change. 2021 Nov; 11(11):952-7.
[8] Albalate D, Bel G, Mazaira-Font FA. Geography and regional economic growth: The high cost of deviating from nature. Journal of Regional Science. 2022 Mar; 62(2):360-88.
[9] Paudel RC. Economic growth in developing countries: Is landlockedness destiny? Economic Papers: A journal of applied economics and policy. 2014 Dec; 33(4):339-61.
[10] Li X, Zhou Y, Zhao M, Zhao X. A harmonized global nighttime light dataset 1992–2018. Scientific data. 2020 Jun 4; 7(1):168.
[11] Cengiz D, Dube A, Lindner A, Zipperer B. The effect of minimum wages on low-wage jobs. The Quarterly Journal of Economics. 2019 Aug 1; 134(3):1405-54.