Where State Presence Backfires: How State Infrastructure Reshapes Protest Geography

Abstract

How does state infrastructure affect protest? We argue that state infrastructure, while designed to extend state presence and integrate peripheral regions, can paradoxically relocate dissent to politically sensitive centers. Leveraging China's ambitious high-speed rail (HSR) expansion as a quasi-natural experiment, we find that HSR mitigated protests in peripheral areas but intensified them in central regions within each province. We build a formal model based on spatial interaction to explain this finding. Empirical results generally support the deterrence effect in peripheral regions where HSR enhances state presence and deters residents from protest participation, and the agglomeration effect where HSR encourages migration towards central areas and intensifies conflicts. Our findings highlight the *authoritarian infrastructure paradox*: while state infrastructure is intended to strengthen economic integration and extend control in peripheral regions, it can inadvertently relocate dissent to politically sensitive urban centers, posing new challenges for regime stability.

Keywords: State infrastructure; Protest; State presence; Political geography.

Word Count: 11,986

1 Introduction

Economic modernization is often accompanied by shifts in political geography (Polanyi et al., 1944; Huntington, 2006). The expansion of state infrastructure, such as railways, not only deepens economic integration (Faber, 2014; Donaldson, 2018), but also projects political control over remote regions, thereby enhancing state capacity (Mann, 1984). In authoritarian regimes in particular, infrastructure is often viewed as a tool for consolidating political control: roads, surveillance devices, and police stations intensify state monitoring and repression (Xu, 2021; Chang and Wang, 2024; González et al., 2025). These systems enable the state to penetrate local society, extract economic resources, and reinforce authoritarian rule (Mann, 1984; Shue, 1990).

However, there remains no clear consensus on how infrastructure impacts the interplay between state power and political resistance. While conventional wisdom asserts that heightened state presence can reduce rebellions (Schuler and Truong, 2020; Chang and Wang, 2024), recent studies provide counterexamples, suggesting that it can, in some cases, provoke violent resistance (Ying, 2021; Chiovelli et al., 2024; González et al., 2025). These conflicting findings suggest that the relationship between infrastructure and protest may not be monotonic. Based on existing studies, this article offers a spatial perspective to reconcile these empirical tensions. We highlight the *authoritarian infrastructure paradox*, where state infrastructure can suppress resistance in remote regions but incite it in central areas, thus reshaping the *political geography* of protest.

The *political geography* matters as protest becomes more dangerous when it occurs closer to the capital. During the French Revolution, the 550,000 Parisian crowd contributed mostly to the regime change despite there being many royalists willing to defend the monarchy in the 29 million Frenchmen (Tilly, 2003). In modern times, successful protests also typically started and upsurged at the center, such as the 2006 “Orange Revolution" in Ukraine, the 2014 protest in the same country, and the 2011 “Jasmine Revolution" in Tunisia. Especially in authoritarian contexts, unrest in capital centers is widely regarded as disproportionately destabilizing (Campante et al., 2019).

Building on these insights, we examine how state infrastructure reshapes the spatial distribution of protest and highlight the *authoritarian infrastructure paradox*. Large-scale, inter-regional

projects extend state presence into previously remote regions but also induce significant population mobility. These spatial dynamics, often overlooked in studies of contention politics, help explain why infrastructure can simultaneously suppress dissent in peripheral areas while intensifying contention at the core. Designed to strengthen economic integration and political control, infrastructural expansion may instead backfire by concentrating protests in politically sensitive capitals, thereby creating new vulnerabilities for authoritarian stability.

To support this argument, we study the world's largest authoritarian state, China, and its ambitious high-speed rail (HSR) program. Several features make China's HSR a compelling case. First, its massive scale is unparalleled. By 2023, Chinese HSR lines span 26,098 miles (the longest in the world), affecting over one billion citizens and serving 1.67 billion passengers annually. Second, its centralized organization distinguishes it from other infrastructure programs, which are often shaped by local governments or societal forces (Olken, 2007; Xu and Yao, 2015). The *Mid-to-Long Term Railway Network Plan* in China established a single, nationwide HSR network, with the Chinese Ministry of Railway (hereafter MOR) funding most of its cost. Third, China's vast geography debilitates the state's ability to extend its power to remote regions (Weber, 2016; Mann, 1984). Transportation infrastructure, thus, becomes indispensable for monitoring local officials, exerting social control, and enhancing state capacity. The wide territory of each province also provides us with sufficient variation to document the heterogeneous impact of HSR construction. Last but not least, examining the issue in China, the world's largest developing country and authoritarian state, could offer valuable insights for other developing nations and non-democracies.

Empirically, we construct a panel dataset by matching prefecture-year-level data on HSR opening with data on protest events in China from 2010 to 2016. Protest events are quantified using the Collective Action from Social Media (CASM) dataset, aggregated to the prefecture-year level (Zhang and Pan, 2019). Using difference-in-differences (DID) as our empirical strategy, our baseline results reveal a spatially heterogeneous effect: HSR openings reduce protests in areas far from the provincial capital (peripheral areas, hereafter) while fostering protests in regions close to the capital (central regions, hereafter). For an average province, HSR opening leads to 0.467 s.d. more (approximately 17.7) protest events in the provincial capital each year. Our dynamic

analysis further shows no significant differences in protest trends between treatment and control groups before HSR opening; moreover, following HSR connectivity, protest events rise over time in central regions while wane gradually in peripheral areas, reflecting sustained effect. The results remain robust when relaxing linearity assumptions, applying heterogeneity-robust estimators (De Chaisemartin and d'Haultfoeuille, 2024), exploiting strike data from the China Labor Bulletin, and controlling for a range of potential confounders.

To explain HSR's spatially heterogeneous effects, we develop a formal model of transportation networks and protest, extending Passarelli and Tabellini (2017), and test its implications empirically. We find evidence for a *deterrence effect*: by expanding state presence in remote regions, HSR suppresses protest participation. At the same time, HSR generates an *agglomeration effect*: migration flows toward provincial centers increase the concentration of low-skilled rural migrants, whose exclusion and discrimination fuel higher protest rates in urban cores. By contrast, we find no support for a *spillover effect*—that peripheral areas benefit economically at the center's expense—or for an *monitoring effect*, whereby stronger oversight of local officials improves governance and reduces dissent. We also rule out alternative explanations, including HSR-related protests, protest commuting, or shifts in public spending.

This study contributes to research on how states expand *state presence* and project power across their territory (Rogowski et al., 2022). Existing research predominantly examines the positive impacts of state presence, like mitigating political violence (Kalyvas, 2006), promoting economic development (Acemoglu, García-Jimeno, and Robinson, 2015; Mattingly, 2016), and reducing monitoring cost (Mastrorocco and Teso, 2023). Recent studies have documented its potential negative effects on social stability (Ying, 2021; Chiovelli et al., 2024; Calamunci and Lonsky, 2025; Blair, 2025). Our study contributes to this emerging discourse by emphasizing the *spatially heterogeneous effects* of extending state presence, proposing the *authoritarian infrastructure paradox*: large-scale, inter-regional projects designed to consolidate regime control can instead backfire, reinforcing state presence in peripheral regions while intensifying protest in politically sensitive urban centers.

We also advance the understanding of the political implications of transportation infrastruc-

ture. Prior research shows that infrastructure can shape electoral outcomes (Voigtländer and Voth, 2025; Nall, 2015; Akbulut-Yuksel, Okoye, and Turan, 2024), affect crime patterns (Calamunci and Lonsky, 2025), condition state repression (González et al., 2025), and facilitate mobilization (Pengl et al., 2025), yet its implications for protest remain less understood. While historical work (Traugott, 2010) and recent empirical studies (Ying, 2021; Müller-Crepon, Hunziker, and Cederman, 2021; Pruett, 2024) suggest that roads can influence conflicts, they largely overlook how such effects vary across space. From the perspective of *spatial politics*, we advance this literature by demonstrating how infrastructural projects may simultaneously suppress dissent in peripheral regions and intensify contention in political centers, elaborating the underlying mechanisms clearly.[1]

Our study also extends the literature on protest geography. Existing research examines how urbanization induced more protests to concentrate in city centers (Wallace, 2014; Beissinger, 2022). [2] Building on Huntington's classic insights (Huntington, 2006), we show how a state-driven modernization shock—an ambitious state infrastructure project such as HSR—can exogenously reshape the geography of protest and generate new challenges for social stability.

Finally, our findings align with a burgeoning literature on authoritarian resilience, which highlights that efforts to strengthen state control can paradoxically backfire. Expanding connectivity can lower coordination costs and spur mobilization, from mobile phones and social media to railroads (Manacorda and Tesei, 2020; Enikolopov, Makarin, and Petrova, 2020; Qin, 2017). Development programs intended to pacify populations may provoke violence (Crost, Felter, and Johnston, 2014; Sexton and Zürcher, 2024). We extend this scholarship by demonstrating the counterproductive effects of HSR and by shifting the focus from *whether* state-building effort triggers unrest to *where* dissent emerges.

1. Bautista et al. (2023) documents heterogeneous repression in Chile shaped by the location of military bases; Schuler and Truong (2020) shows that the Internet's effects on protest vary by rural–urban context; Cao and Chen (2022) document rising social conflicts during canal abolishment in ancient China. We differ by focusing on the interaction between economic centripetal forces and repressive state power, producing a spatially heterogeneous pattern of protest.

2. Existing studies on China's protests are primarily qualitative, focusing on protest structure ("rightful resistance" (O'Brien and Li, 2006; Cai, 2008), "atomize protest" (Rho, 2023)), protests in different regions (Lee, 2007), protest motives (Cai, 2010; Lee and Zhang, 2013), the way that the state managed protests (Fu, 2017; Li, 2019; Elfstrom, 2021), and new patterns of protest (Steinhardt and Wu, 2016; Steinhardt and Yang, 2025). Our study is among the first to explore the geography of protest patterns.

2 Where State Infrastructure Backfires: Political Geography of Protest

State infrastructural power is often regarded as a hallmark of the modern state (Mann, 1984). It is through infrastructure development that states can exert control over even the most remote regions (Weber, 2016; Mann, 1984). For instance, the Indian government initiated the railway project in 1853, which connected India's five major provincial capitals along direct routes (Donaldson, 2018). Similarly, China's HSR program connected provincial capitals with local prefectures. Yet state infrastructure does not always serve to facilitate political control over local areas. In this section, we theorize an *authoritarian infrastructure paradox* that elucidates why, at times, the development of state infrastructure, especially railways, leads to political control and suppression in remote regions while simultaneously increasing challenges to state authority and instances of protest in urban centers.

2.1 Infrastructure as State Presence: Extending Control and Deterring Protest

State infrastructure has the potential to reshape the political landscape. A few studies have investigated how infrastructure influences election outcomes, either signaling the ruling party's visibility and competence (Voigtländer and Voth, 2025; Akbulut-Yuksel et al., 2024) or driving the flow of different groups of voters (Nall, 2015). More importantly, the state exerts its authority in part through the establishment of its infrastructure. Transportation infrastructure not only signifies state presence, in the sense that it represents the state's authority in constructing large-scale public facilities (similar to public bureaus, shown by Chang and Wang (2024)), but also improves state presence, as it directly reduces the time (and transportation cost) for state forces to reach peripheral regions (González et al., 2025). By enhancing the reach and efficacy of state mechanisms, infrastructure can act as a deterrent to collective actions that oppose state interests.

For one thing, the state can enhance governance efficiency by projecting its authority and exerting control over local officials, thereby ensuring compliance, summarized as the *monitoring effect*. For an inspector choosing the optimal city for anti-corruption, her preference is shaped by local corruption rate and the cost to reach there. Naturally, she is more inclined to inspect regions

with more severer corruption and less transportation cost. For local government officials, they may corrupt so long as not being inspected. In equilibrium, the more remote officials become more corrupt (distance advantage). After HSR opening, such distance advantage is reduced for remote areas, which also reduces the equilibrium corruption rate. As residents generally detest corruption, reduction in corruption helps lower their aggrievement and reduce the protest rate.[3]

Hypothesis 1 (Monitoring Effect) *HSR opening can facilitate investigations into local officials, alleviating corruption and reducing protests in remote areas.*

The enhanced state presence not only inflicts local officials, but pressures local residents, raising their perceived costs to participate in protests, insurgencies, or revolutions, summarized as the *deterrence effect*. For a public security leader choosing the city for anti-protest, her preference is also shaped by local protest rate and the difficulty in reaching there. By a similar argument as before, there is a distance advantage: in equilibrium, residents in more remote areas are more inclined to protest, as they are less likely to be caught. After HSR opening, however, such advantage is reduced for remote areas, which give rise to the repression rate (i.e., the cost of protest). In other words, residents are deterred by increasing state presence. For fear of being caught to jail, they would be more prudent and the protest rate is reduced.

Hypothesis 2 (Deterrence Effect) *HSR opening can extend state presence, deterring residents from launching protests, thus reducing protests in remote areas.*

2.2 Where Infrastructure Backfires: Protest Concentration in Urban Centers

While the aforementioned analysis elucidates the beneficial aspects of state infrastructure for local control and the suppression of protests, it overlooks the potential for large-scale, inter-regional infrastructure like railways to reshape local economies spatially. The mass population mobility facilitated by railways may draw peripheral populations into provincial capitals, thereby providing fertile ground for protest movements to take root and flourish.

3. See Seligson (2002) and Anderson and Tverdova (2003) for relation between corruption and regime support.

On the one hand, HSR is more economical than traditional transport means (like buses or trains) when the distance is far. Hence, HSR opening helps to reduce local residents' transportation cost. In a typical center-periphery economy[4], people are inclined to migrate to the center for better job opportunities. The reduction in transportation cost, thus, directly foster such centripetal migration, leading to population agglomeration, summarized as the *agglomeration effect* (Faber, 2014; Dong, 2018; Dong et al., 2020).

More population does not necessarily mean more protests. However, the migrants from peripheral regions to more developed central regions are mostly agriculture-to-non-agriculture migrants (A-to-NA migrants, hereafter) who usually possess low skills and poor education, thus living a rather distressed life. In the context of China, the proportion of A-to-NA migrants to all migrants is at least 76% in 2000-2020.[5] Moreover, regulations (like the *hukou* system in China) may prevent migrants from getting equal rights with urban residents, such as in children's education, parents' healthcare, and housing for themselves. What's more, most of them overwork for long, with poor living conditions, a few of whom suffer from wage arrears that incur severe economic vulnerability. Summed up, these factors propel their protest willingness by creating such "relative deprivation" (Gurr, 1970). What's more, with more "potential protesters", protests are more likely to succeed as the probability that a critical threshold of participators is reached such that the crowd overwhelms the police becomes higher (Morris and Shin, 2000), or equivalently, the probability of being arrested is smaller in a large crowd (Le Bon, 1897). In general, the centripetal migration give rise to protests in central areas but reduce them in remote areas.

Hypothesis 3 (Agglomeration Effect) *HSR opening can foster migration, thus*

- *increasing protests in central areas by attracting more migrants;*
- *reducing protests in remote areas by pushing residents away.*

On the other hand, HSR may foster the overflow of resources and industrial activities, as firms can exploit more resources at a lower cost in remote areas by establishing subsidiaries (Bernard

4. This incorporates a wide range of settings, like the monocentric model (Duranton and Puga, 2015; Wang et al., 2020). In reality, Chinese provinces, states in the U.S., and many other sub-national districts follow similar patterns.
5. See Table 1 of Tombe and Zhu (2019) or China's seventh national census for details.

et al., 2019; Gumpert et al., 2022). More importantly, individuals should be indifferent between migration or staying in their hometown in the spatial equilibrium. Thus, there should be an increase in wage (or decrease in living costs) to offset the decrease in transportation cost, with a larger rise/decline in more distant regions (Charnoz et al., 2018; Lin, 2017; Baum-Snow et al., 2017). In equilibrium, residents living at remote areas should enjoy higher wages or lower living costs, thus becoming less inclined to protest, whereas those in central regions develop higher protest willingness suffering from higher living costs and lower wages.

Hypothesis 4 (Spillover Effect) *HSR opening can foster resource spillover, thus*

- *increasing protests for central areas with lower wages or higher living costs;*
- *reducing protests in remote areas with higher wages or lower living costs.*

To sum up, the four hypotheses shed light on different meanings of HSR opening, yet they are consistent that there would be rising protests in the center but less in the periphery, summarized in Hypothesis 5.

Hypothesis 5 (Main Prediction) *HSR opening would give rise to protest events in central regions while reduce protests in peripheral areas.*

In sum, our theory highlights that state infrastructure reshapes the *spatial distribution* of protest rather than simply increasing or decreasing its overall frequency. This spatial reconfiguration is politically consequential because, especially in authoritarian contexts, unrest in capital cities is disproportionately destabilizing (Campante et al., 2019). Concentrating dissent in provincial capitals may generate far greater challenges for regime stability than reducing protest in peripheral regions. This dynamic captures the essence of the *authoritarian infrastructure paradox*: while state infrastructure is intended to strengthen economic integration and extend control in peripheral areas, it can inadvertently relocate dissent to politically sensitive urban centers, thereby posing new threats to authoritarian rule.

2.3 Mapping Argument to Formal Model

Building on Passarelli and Tabellini (2017), we present a simple theoretical framework in which transportation infrastructure shapes protest participation through four mechanisms: enhanced monitoring of local officials that improves governance (*monitoring effect*); deterrence of residents by strengthening state presence (*deterrence effect*); migration flows that concentrate potential dissenters in urban centers (*agglomeration effect*); and the redistribution of economic activity across space (*spillover effect*). Together, these mechanisms determine how infrastructure reshapes the spatial distribution of protest. The framework provides micro-level foundations for our empirical analysis and generates testable predictions that allow us to disentangle competing explanations. For the full model setup, derivations, and technical details, see Appendix A.

3 Background

3.1 China's High-speed Rail (HSR) System

HSR, also known as bullet trains, operates at a speed of 250 km per hour (kmph) or higher in China, in contrast to conventional trains that typically operate at speeds not exceeding 150 kmph (Zheng and Kahn, 2013), despite several rounds of speed acceleration. The *Mid-to-Long Term Railway Network Plan*, ratified by the Chinese State Council in 2004 and subsequently updated in 2008, was largely completed by 2016 (Dong, 2018).[6] The plan emphasizes enhancing connectivity between provincial capitals and local prefectures. By 2023, China has constructed 159 HSR lines spanning a total mileage of 29,826 miles (48,000 km), with 1.67 billion passengers annually.[7] The budget for HSR construction has grown from $14 billion in 2004 to a substantial $719 billion by 2022, taking up a significant portion of the government's budget.[8]

HSR remains a favored mode of transport due to its speed, reliability, and comfort, delivering a

6. The plan outlines the direction of HSR development. For details, see https://www.gov.cn/xinwen/2016-07/20/5093165/files/1ebe946db2aa47248b799a1deed88144.pdf. Check the updated content at https://www.nra.gov.cn/jglz/ghyj/202104/t20210428_199680.shtml.

7. See https://www.gov.cn/yaowen/liebiao/202501/content_6995868.htm.

8. Sourced from https://www.gov.cn/xinwen/2023-01/04/content_5734924.htm.

competitive advantage over airlines and trains (buses) for mid-length journeys (Amos et al., 2010). The expansion of HSR has notably enhanced intercity transportation, particularly linking remote areas with central megacities (provincial capitals) (Dong et al., 2020). While planned purely for economic purpose, the HSR network may have strong political implications, as we detail below.

3.2 The Political Economy of HSR

The Chinese administrative hierarchy is structured into four levels, as shown in Figure B1. The state comprises 22 provinces, 5 ethnic autonomous regions, and 4 municipalities. We focus on provinces and ethnic autonomous regions, both of which are subdivided into prefectures, while municipalities are cities with higher administrative level, not consisting of prefectures. Figure B2 presents the spatial distribution of Chinese provinces and their capitals. We focus on the political economy within one province, where there is a center (the provincial capital) and some prefecture-level cities and autonomous prefectures (both called "prefectures" hereafter). Figure B3 exemplifies the pattern using Sichuan Province. Central cities are Chengdu and its adjacent prefectures, while peripheral regions refer to those far from Chengdu (usually near the boarder).

Political Geography within Provinces Three main factors impel us to focus on political geography within provinces. The first reason stems from the unique economic geography of China, where political power and economic activities are heavily centralized in provincial capitals. Provincial capitals are, by definition, political centers. They also possess significantly more resources that attract migrants from other prefectures within the province. As is shown in Figure B1, the average population of provincial capitals exceeds that of prefecture-level cities and autonomous prefectures by 2.5 and 4.5 times, respectively. Economic agglomeration is more pronounced, with 23 provinces' primacy of provincial capital (ratio of capital's GDP to provincial GDP) exceeding 20%, including 10 above 30%, and 2 over 50%. Hence, the political center overlaps with the economic center in most provinces.[9] This phenomenon does not apply to the country as a whole, with Beijing and Shanghai serving as the political and economic centers, respectively. In addition, China's

9. This assumption does not restrict our main finding, as shown in Table D7 in the Online Appendix.

large territory and household registration system (*hukou*) impede inter-provincial migration.[10] As a result, political power and economic forces often operate separately at the national level.

The second reason is rooted in China's political structure. The central government retains control over military, diplomatic, and ideological matters to ensure national unity, while delegating other powers to the province level. Provinces enjoy autonomy over key areas, including legislation (via provincial congresses), judicial decisions (by provincial courts and procuratorates), personnel appointments (by provincial party committees), development plans (by provincial government, specifically the development and reform commission), a system summarized by Xu (2011) as "regionally decentralized authoritarianism" (RDA, hereafter). In addition, Chinese provinces are large in both territory (351,481km^2 on average) and population (54.9 million on average), comparable to mid-sized European countries. Thus, each province functions almost as a "state", with the provincial capital exerting substantial political influence on other prefectures. While local governments (prefecture/county/town-level) hold some authority, their power is more restricted, and their territory is too small to have significant internal variation. In sum, the RDA structure makes it appropriate to focus on provinces as the primary political entities.

A third relevant factor is the process of HSR construction, which involves collaboration between central and provincial governments. While the railway center[11] designs the HSR lines, their actual construction relies on economic resources from the provincial government. To address local needs, provinces also conduct local HSR plans seeking central approval. Take "the 14th Five-year Plan" as an example. The railway center announces a grand HSR project at the state level, subdivided by many small projects for each province.[12] Many of the newly constructed lines, in particular, serve provinces' needs to strengthen economic ties and extend political control to peripheral areas. As shown in Figure B4, HSR openings in other prefectures always occur later than their provincial capitals, so HSR opening directly signals connectivity to the capital.

Local governments, however, possess minimal decision-making authority during the route-

10. Intra-provincial migration dominates the general migration pattern in China (approximately 55%). See Table 1 of Tombe and Zhu (2019) or https://www.thepaper.cn/newsDetail_forward_12902468 for details.

11. This refers to the MOR before 2013. To separate government functions from corporate functions, the MOR is split into the China Railway Limited and the National Railway Administration (affiliated to the Ministry of Transport).

12. See https://www.qianzhan.com/analyst/detail/220/211125-230ef9b5.html.

planning process. Local officials merely facilitate construction conditions, having a negligible impact on the already decided HSR lines. Thus, HSR construction affects but is seldom affected by, local governments, providing an ideal setting to explore the implications of state infrastructure.[13]

Geographical Distance The nature of HSR makes geographical distance a pivotal factor. HSR possesses competitive advantage over airlines for journeys shorter than 500 miles, whereas traditional means like trains or buses are considered more suitable than HSR within 100 miles (Pan, Shroff, and Zhang, 2023). Thus, HSR matters particularly for journeys between 100 and 500 miles, which also justifies our primary focus on within-province outcomes (rather than within-country or within-prefecture ones). As shown in Figure B5, over 80% of prefectures lie within 100-300 miles from provincial capitals, making them heavily influenced by HSR.

4 Data and Empirical Strategies

4.1 Dataset

By matching data on HSR construction with protest event data in China from 2010 to 2016, we examine the heterogeneous impact of HSR opening on protests at varied geographical distances from provincial capitals. Our baseline primarily utilizes the dataset at the prefecture-year level. We exclude samples from municipalities and cities without available information, eventually obtaining a prefecture-year level dataset covering 279 prefectures.

4.1.1 Data on protest events

Collective Action from Social Media Dataset. Traditional data sources for protest events, such as newspapers and press releases, tend to overrepresent larger-scale and more sensational protests (Chang and Wang, 2024). This issue becomes particularly challenging in authoritarian states with strict media censorship, where traditional news media cannot report protest events.

13. Considering that particular local officials may influence HSR opening (possibly through non-procedural way, summarized by Ma (2022) as "localized bargaining"), we conduct robustness checks on officials' personal characteristics in the Appendix. Even if they might hold some power, it hardly affects our results.

To address the issue, we use the Collective Action from Social Media (CASM) dataset, recently made available by Zhang and Pan (2019). This dataset employs convolutional neural networks on image data and recurrent neural networks on text data in a two-stage classifier to identify social media posts about offline protest events. The CASM dataset identifies more than 100,000 protest events in China from January 1, 2010 to June 30, 2017, sourced from Sina Weibo, China's biggest micro-blogging platform. By 2016, Weibo had 297 million monthly active users, accounting for 41.83% of Chinese Internet users. We aggregate this event-level dataset to prefecture-year level.

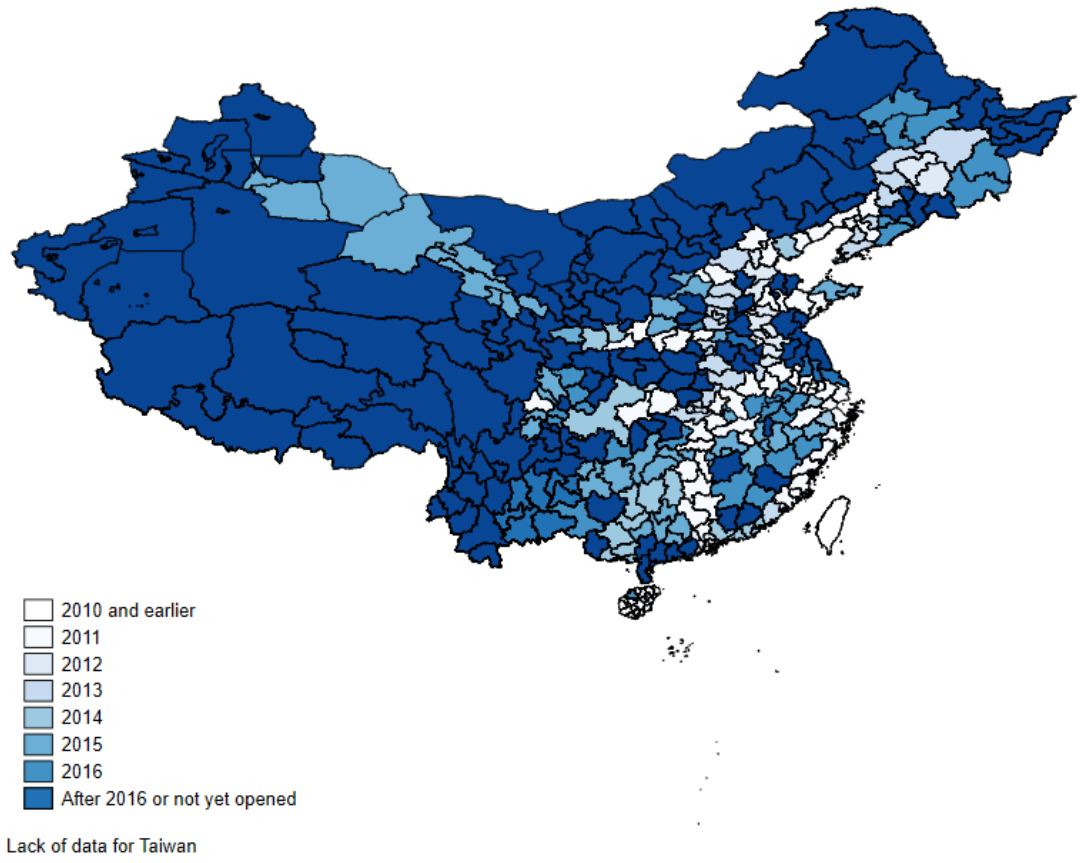
Data from social media, while comprehensive, may introduce regional biases if media usage is correlated with the timing of HSR opening. We take two approaches to addressing this issue. First, Table B1 shows that HSR does not increase social media penetration or total Weibo posts, so even if there is potential regional biases in media coverage, it cannot confound our estimation. Second, we use alternative data sources that do not rely on social media to complement CASM.

China Labor Bulletin. We use the strike dataset for supplementary validation. China Labor Bulletin (CLB) is a non-profit organization based outside mainland China, collecting data on workplace strikes from various media sources. The dataset covers strikes in different industries, enabling us to explore which groups of workers are most likely to engage in protests. Unlike the CASM dataset, CLB data are not aggregated from social media, thus avoiding potential biases related to regional differences in internet penetration. Additionally, CLB provides a longer post-HSR opening sample compared to CASM (2010-2018).[14] Table B2 presents descriptive statistics for our main variables in the CLB dataset. As shown in Figure B6, the correlations between protest counts in the CASM, CLB, and GDELT datasets are strongly positive, so they should provide consistent results in our subsequent analysis (with different meanings).

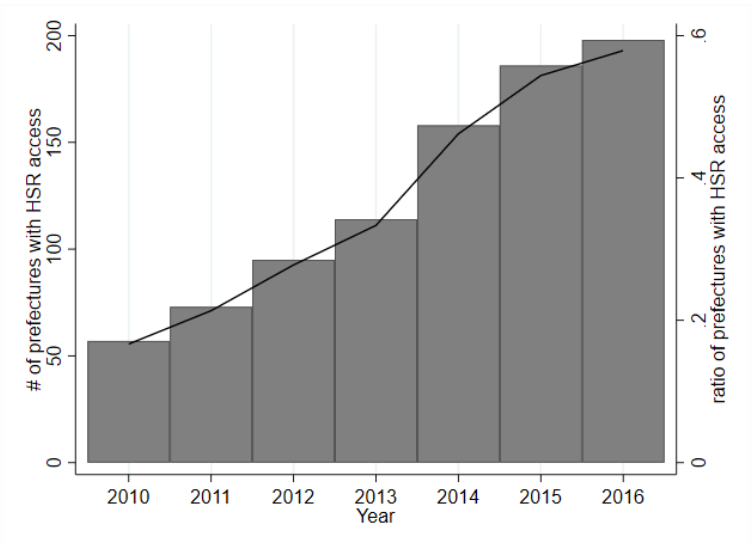
4.1.2 Data on HSR

We obtain the data on whether the prefecture has access to HSR in a certain year from the China Railway Yearbook, encompassing details such as construction commencement dates, opening

14. Qin et al. (2024) employ the dataset to show the spatial contagion of strikes in China. Campante et al. (2023) also use this dataset to study the effects of export slow-down on labor unrest in China.



(A) **Distribution of HSR Opening Times Across Prefectures.**



(B) **Trends in Cities with HSR Connectivity.**

FIGURE 1 HSR Openings.

Notes: This figure illustrates the distribution of HSR openings across different regions and time periods in China. Panel A shows the variation among different prefectures, while Panel B provides an overview of the number and proportion of HSR-opening cities across different years.

dates, route lengths, designated speeds, and ridership. We cross-verify the stoppage points along each operational line using official data from the railway service website (https://www.12306.cn/). Further, we utilize the data recently made available by Davis, Qian, and Zeng (2025) on China's railway and highway network to calculate changes in market access. Figure 1 shows the opening times of different prefectures. Notably, in our sample, the number of prefectures connected to the HSR network increased from 53 in 2010 to 180 in 2016, covering approximately 60% of all prefectures in China.

4.1.3 Control variables and other datasets

China City-specific Characteristics. To control for regional influences, we incorporate predetermined prefectural characteristics from the *China City Statistical Yearbook*, including the proportion of the primary sector, that of the secondary sector, per capita GDP, population, and urban area in 2010. Also, we manually collected detailed data on Chinese officials to rule out alternative explanations, and employ various datasets to explore the mechanisms, as described below.

Individual Surveys. We employ a series of individual-level micro surveys to explore the mechanisms in detail. First, we utilize the China Family Panel Studies (CFPS), a nationally representative social survey, to examine how HSR openings affect individuals' emotions and political attitudes. Second, we use the Ideology Survey of Chinese Internet Users (ISCIU) dataset obtained from Ma and Wang (2015), which provides data on individual political expressions. Third, we employ the China Migrants Dynamic Survey (CMDS), an annual cross-sectional survey covering a large sample of migrants from 267 prefectures (642,700 individual-years). The CMDS offers unique information on the status of migrant workers in central cities.

4.2 Descriptive statistics

Table 1 presents the summary statistics for our sample, which comprises 279 prefectures annually. On average, 29 protest events take place each year in a prefecture.

Based on the raw data, we compare protest geography before and after HSR opening in Fig-

TABLE 1 Summary Statistics

	(1)	(2)	(3)	(4)	(5)
Variable	Obs	Mean	Std.Dev.	Min	Max
Panel A: Independent Variables					
HSR	1,896	0.412	0.492	0	1
Distance	1,896	130.992	72.897	0	301.709
lndistance	1,896	4.425	1.483	0	5.713
Panel B: Prefecture-year Outcomes[a]					
Protest	1,896	28.790	37.910	0	374
Protestsd	1,896	0	1	-0.759	9.105
Panel C: Predetermined Control at 2010					
Primary industry(%)	279	3.641	7.656	0.030	56.180
Secondary industry(%)	279	42.456	13.980	7.240	81.850
GDP per capital(yuan)	279	33,266.47	22,771.95	5,304	175,125
Population (per 10,000 people)	279	440.880	310.639	21.8	3,303.45
Area[b]	279	9.363	0.818	7.015	12.442

Notes: Table 1 presents descriptive statistics for our main variables.
[a]: $Protest$ is the number of protests, while $Protestsd$ is the standardized number.
[b]: We define $Area$ as the logarithm of the prefecture's land area (unit: square kilometers).

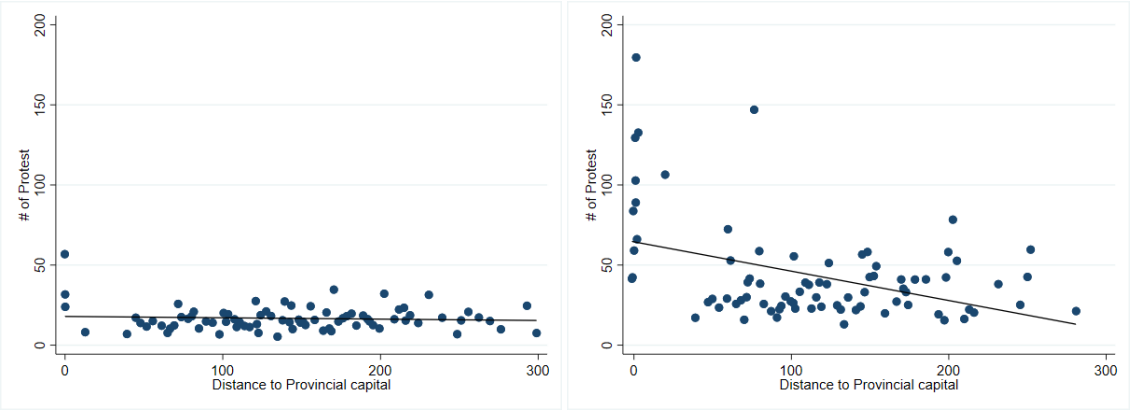
ure 2. Protests are rather sporadic before HSR opening, after which they show a declining trend with geographical distance. This suggests that the HSR opening may reshape the political geography by redirecting protests to the center. Below we use the Difference-in-Differences (DID) method for causal identification.

4.3 Empirical Strategy

We begin by applying the DID approach, comparing the relative changes in protest events in prefectures with HSR opening to those without. The model specification takes the following form:

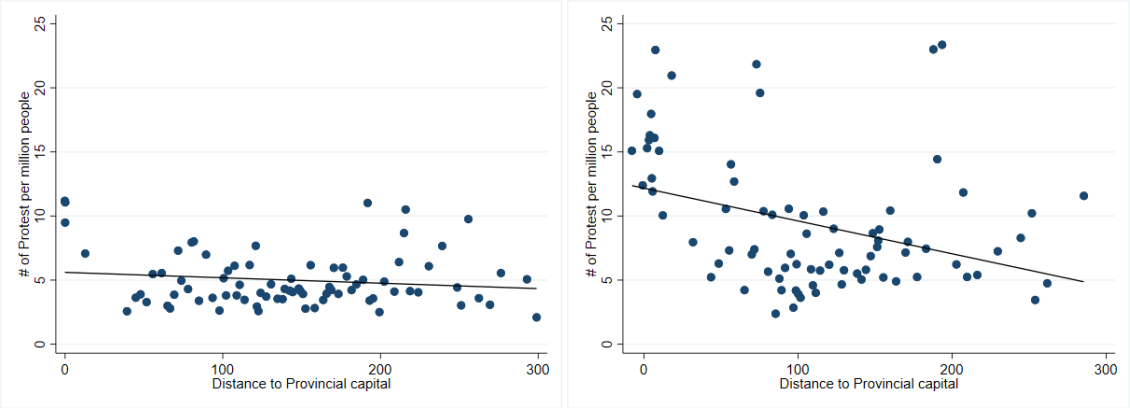
$$Protestsd_{cpt} = \alpha_0 + \alpha_1 HSR_{cpt} + \alpha_2 \phi' \boldsymbol{X_c} \times T_t + \lambda_c + \theta_{pt} + \epsilon_{cpt}, \quad (1)$$

where, $Protestsd_{cpt}$ represents the number of protests in province p, prefecture c, year t, standardized for ease of interpretation. HSR_{cpt} is a dummy variable that takes the value 1 if a prefecture has HSR access and 0 otherwise. λ_c is the prefecture fixed effect that allows us to identify the



(A) **The geographical distribution of protest events before HSR opening.**

(B) **The geographical distribution of protest events after HSR opening.**



(C) **The geographical distribution of protest events per million people before HSR opening.**

(D) **The geographical distribution of protest events per million people after HSR opening.**

FIGURE 2 The geographical distribution of protest events before and after HSR opening.

Notes: This figure illustrates the geographical distribution of protest events before and after HSR opening. In Panel A-B, the y-axis displays the normalized number of protest events for each prefecture, while in Panel C-D, the y-axis shows the number of protest events per million people. The x-axis in all panels represents the geographical distance from the prefecture to the provincial capital. Panels A and C illustrate the spatial distribution pattern of protests before HSR opening, while Panels B and D depict that after HSR opening.

impact from within-prefecture variations, absorbing time-invariant local factors, like historical resistance traditions. θ_{pt} is the province-year fixed effect that excludes potential time-varying confounders at the province level, such as changes in macro-level policies. We also control predetermined features $\boldsymbol{X_c}$ (sector shares, per capita GDP, population, and urban area in 2010) interacted with time dummy T_t to partial out their potential influence on protests. For each regression, we report two sets of standard errors. First, the standard errors clustered at the prefecture level, as our dependent variable HSR varies between prefectures. Second, we report the Conley standard errors following Conley (1999), as commonly used to account for potential spatial correlations.

This identification relies on two sources of variation: differences in HSR openings across cities and the pre- and post-HSR changes in protests within the same city. Thus, the key parameter of interest, α_1, captures the average treatment effect on the treated (ATT) of HSR opening on protests. However, as predicted by the theory in Section 2, HSR openings may influence the geographic distribution of protests, with varying effects on central regions and peripheral areas. Consequently, the total effect of α_1 may be offset by both positive and negative effects across cities, potentially resulting in an overall null effect on protests. To explore this further, we introduce a prefecture-to-capital distance variable to examine how geographical proximity moderates the effect of HSR openings on protests:

$$Protestsd_{cpt} = \beta_0 + \beta_1 HSR_{cpt} + \beta_2 HSR_{cpt} \times lndistance_c + \beta_3 \phi' \boldsymbol{X_c} \times T_t + \lambda_c + \theta_{pt} + \epsilon_{cpt}, \quad (2)$$

The only difference between Equation 1 and Equation 2 is the inclusion of the interaction term between distance to the provincial capital ($lndistance_c$) and HSR.[15] The single distance term is absorbed by the prefecture fixed effect λ_c. The parameters of primary interest in Equation 2 are β_1 and β_2. β_2 indicates the change in protests as the distance from the provincial capital increases after HSR opening, while β_1 represents the change of protests in the provincial capital.

15. We employ the log form following extant studies (See e.g., Glaeser et al. (2012); Campante et al. (2019)), and plus 1 to account for zeros. Our findings are unchanged when using alternative measures, including the logged distance without the adjustment or the raw distance.

Identification Assumptions Our identification does not require the exogeneity of HSR opening. Instead, what is required is the parallel-trend assumption: in prefectures without HSR, the trend of protests over time and with respect to distances from the provincial capital is similar to those with HSR openings. We enhance the plausibility of the assumption in two ways. First, we use a specification that includes province-year fixed effects. Second, we estimate a fully dynamic version to test potential pre-trends, which also allows us to examine the temporal variation of its effect on protests.

The two-way fixed effects (TWFE) estimator may be biased when there are heterogeneous treatment effects over time or across subgroups (De Chaisemartin and d'Haultfoeuille, 2020). This issue is particularly relevant in our context, as we focus on spatially heterogeneous effects. To shed light on heterogeneity by distance, we divide our sample into four subgroups based on their geographical distance to the center (closet, near, mid, and far) and then estimate their the average treatment effects (ATEs) respectively, with consistent results using both TWFE and heterogeneity-robust estimators.[16] This also helps to relax the linear assumption in Equation 2.

Our identification strategy also assumes a no anticipation effect, where individuals do not preemptively alter their protest decisions in anticipation of HSR opening. To address non-random expectations and other local confounders, we employ the recentered-IV approach proposed by Borusyak and Hull (2023). Following their work, we also use market access as an alternative measure of HSR exposure. This measure captures the varying qualities of HSR lines and provides additional insights into HSR's heterogeneous effects.

16. Recent studies have proposed a number of new estimators to overcome this issue (De Chaisemartin and d'Haultfoeuille, 2020; De Chaisemartin and d'Haultfoeuille, 2024; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; Borusyak et al., 2024), yet little is known about the way to estimate the staggered DID with an interaction term. We thank Clément de Chaisemartin for recommending us to estimate the effect separately on different groups and employ their hetero-robust estimator. Moreover, our findings are robust to other estimators like CSDID estimator.

5 Empirical Results

5.1 Baseline Estimation

We begin by estimating the direct effect of HSR on protests (Equation 1). Columns (1)-(3) in Table 2 show that HSR opening has a small and statistically insignificant effect in the full sample.This motivates our analysis in Equation 2, focusing on the heterogeneous effects of HSR opening by distance to the provincial capital. In columns (4)-(6), we find a significant increase in protests in provincial capitals following HSR opening, with the effect gradually diminishing as the distance from the capital increases. We control prefecture and year fixed effect in column (4), add province-year fixed effect in column (5), and further employ predetermined-year control in column (6), verifying the robustness of our findings. Overall, we verify the main prediction in Hypothesis 5: on average, HSR opening did not lead to more protests, but it played a pivotal role in reshaping the *political geography* of protests: increasing in central regions while reducing in peripheral areas.

Specifically, after HSR opening, protests in central regions increased by 0.467 standard deviations (s.d.), which diminishes gradually (0.109 s.d.) as the distance from the provincial capital grows. This translates to 17.71 (0.467×37.91) additional protests in the center and 4.13 (0.109×37.91) fewer protests as distance increases. Figure C1 shows the heterogeneous effect by distance explicitly: HSR gives rise to protest events in central regions (positive intercept), but the effect turns negative gradually as the city becomes farther from the center (negative slope).

One potential concern is different pre-trends in protests for different groups. We estimate a fully dynamic version of Equation 2 and test for potential pre-trends in Figure 3. Before HSR opening, there is little difference in protests for the treatment (with HSR) and control (without HSR) groups.[17] Afterwards, protests soar in central regions (plotted in Panel A and B), but diminish with the geographical distance (plotted in Panel C and D). The results suggest that the trend of protests is similar for the treatment and control group. Moreover, the effect persists for long, so the observed effects can hardly be driven by an immediate political spotlight effect (i.e., citizens

17. Before treatment, protest grows a bit with distance, which may reflect lower state presence and control in more remote regions. However, the coefficients are not statistically significant, with little threat to our results.

using the HSR opening as an opportunity to express grievances).

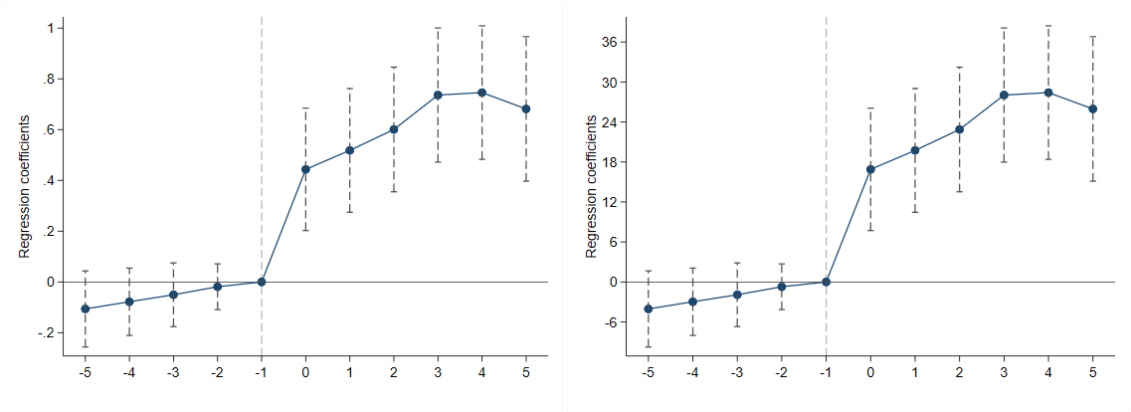
TABLE 2 Baseline Results

Dependent variable	*# of Protest (std)*						*Protest / million*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HSR	0.023	0.005	-0.022	0.474***	0.771***	0.467***	4.364***
	(0.052)	(0.056)	(0.045)	(0.170)	(0.138)	(0.134)	(1.124)
	[0.066]	[0.063]	[0.040]	[0.160]	[0.163]	[0.126]	[1.077]
HSR×*lndistance*				-0.105***	-0.172***	-0.109***	-0.914***
				(0.035)	(0.030)	(0.027)	(0.227)
				[0.032]	[0.033]	[0.025]	[0.200]
Predetermined-year Control			✓			✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓	✓
Year FE	✓			✓			
Province-year FE		✓	✓		✓	✓	✓
Observations	1,994	1,984	1,910	1,957	1,943	1,896	1,896
R-squared	0.785	0.834	0.897	0.786	0.837	0.900	0.846

Notes: Table 2 reports the baseline regression results. The dependent variable in Column (1)-(6) is the standardized number of protests at the prefectural level (*Protestsd*). Column (7) divides the variables by the total population (in millions) (*Protest / million*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. The predetermined-year control variables include the interaction terms between the year dummy variables and the proportions of the primary sector, secondary sector, per capita GDP, population, and urban area in 2010. Robust standard errors in parentheses are clustered at the prefecture levels; those in brackets are Conley standard errors robust for spatial correlation, assuming a cutoff window of 500 km and a serial correlation of 6 years. ***p< 0.01, **p <0.05, *p < 0.1.

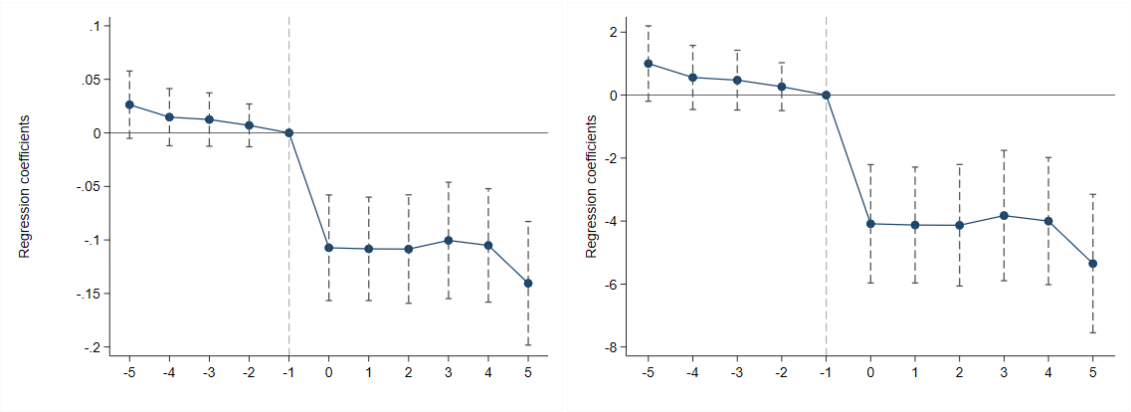
To relax the linearity assumption in Equation 2, we use dummy variables for different distance sections to analyze the respective effect on distinct regions. Table C1 shows a strong positive effect for central regions and a more negative effect for distant prefectures, with consistent evidence using different distance classification. To account for heterogeneous treatment effects, we further address problems with the TWFE estimator using the DID_M estimator proposed by De Chaisemartin and d'Haultfoeuille (2020). Table C2 shows similar estimates: HSR can increase protests in regions closest to the center but reduce them in more distant regions. Additionally, we test for the dynamic effects using the DID_l estimator (De Chaisemartin and d'Haultfoeuille, 2024), which compares switchers and not-yet-switchers.[18] Figure 4 shows no significant pre-trends, and

18. DID_l allows for heterogeneous ATEs and inter-temporal effects. For binary and staggered setting, it is the same as the group-time ATE (CSDID) proposed by Callaway and Sant'Anna (2021). Similar evidence can be drawn by using alternative estimators (e.g., Callaway and Sant'Anna (2021); Sun and Abraham (2021)).



(A) **Pre-trend tests for *HSR* term using the standardized number of protests.**

(B) **Pre-trend tests for *HSR* term using the number of protests.**



(C) **Pre-trend tests for interaction term using the standardized number of protests.**

(D) **Pre-trend tests for interaction term using the number of protests.**

FIGURE 3 Pre-trend tests using prefecture-year level data.

Notes: This figure plots pre-trend tests. The dependent variable is the standardized number of protests at the prefecture level (*Protestsd*) in Panel A and Panel C, while we use the actual number of protests (*Protest*) in Panel B and Panel D. Panel A and Panel B test for single *HSR* term. Panel C and Panel D test for the interaction term *HSR*×*lndistance* in equation (2); All panels control for Province-year, predetermined-year, and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.

more importantly, the effect is significantly positive for the closest group, negative for the farthest group, while insignificant for intermediate groups.

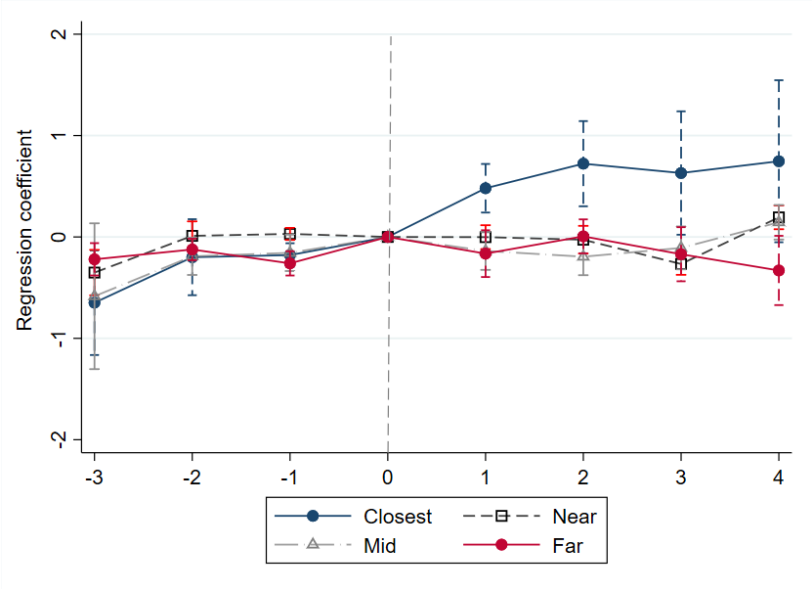


FIGURE 4 Event study in different groups with different distance to provincial capital.

Notes: This figure presents the changes in the frequency of protests (standardized) for different types of prefectures (divided into four groups based on their distance to the provincial capital, from closest to farthest) using De Chaisemartin and d'Haultfoeuille (2024)'s estimator. The dependent variables are the standardized number of protests at the prefecture-year level (*Protestsd*). The dotted line shows a 95% confidence interval.

5.2 Further Classification

Protests vary both in their causes (content) and intensity (form). We use the 11 classifications of protests from the CASM dataset to explore how HSR opening influences specific types of protests. The results are shown in Table 3: HSR opening primarily influences four kinds of protests (those related to unpaid wages, housing issues, education issues, and medical issues), intensifying them in central regions while reducing them in peripheral regions. These four economic issues imply that A-to-NA migrants play a crucial role in our observed effects, since they suffer particularly from wage, housing, healthcare, and education issues (see section 3). We further plot pre-trend tests for these four kinds of protests in Figure C5 and Figure C6. The effect persists for long after HSR opening, suggesting that these four categories contribute significantly to the overall baseline findings. We explore the role of A-to-NA migrants in greater detail in section 6.

The intensity of protests also differs. We examine whether residents engage in drastic or moderate protests after HSR opening. Following Zhang and Pan (2019), we classify them into three forms: conventional (street marches, strikes, public gatherings, demonstrations, and group petitions), disruptive (occupation of buildings, construction of barricades, and cutting off power supplies), and violent (physical conflicts with police or government officials). The first category is rather moderate, while the latter two are drastic. Within the latter two forms, disruptive protests are highly organized and usually well-planned, whereas violent ones tend to be more spontaneous, simply for protesters to vent their grievances. The results are shown in Table C3. All three forms of protest show similar trends, but disruptive ones are affected less by HSR opening. The pre-trend tests in Figure C2 and Figure C3 show that conventional and violent events increase in central regions and decline in remote areas. These findings echo previous results on protest causes. Since economic issues such as wages, housing, education, and healthcare are the primary drivers of protests, it is unlikely that protesters will organize large-scale, coordinated resistance.[19]

19. Previous studies have shown that politically related protests are better-organized (Cantoni et al., 2024).

TABLE 3 Classification Based on Protest Contents

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	All	Unpaid wages	Homeowner	Land	Education	Medicine
HSR	0.467***	0.779***	0.733***	0.283	0.423**	0.602***
	(0.134)	(0.198)	(0.209)	(0.272)	(0.179)	(0.215)
	[0.126]	[0.185]	[0.186]	[0.252]	[0.181]	[0.602]
HSR×*lndistance*	-0.109***	-0.166***	-0.171***	-0.082	-0.109***	-0.129***
	(0.027)	(0.040)	(0.042)	(0.056)	(0.037)	(0.045)
	[0.025]	[0.039]	[0.037]	[0.054]	[0.035]	[0.047]
Observations	1,896	1,896	1,896	1,896	1,896	1,896
R-squared	0.900	0.823	0.821	0.785	0.716	0.732
Dependent variable	Taxi	Fraud	Environment	Pension	Ethnic	Veteran
HSR	-0.010	0.843*	-0.169	0.167	-0.098	-0.048
	(0.179)	(0.454)	(0.260)	(0.146)	(0.289)	(0.214)
	[0.185]	[0.431]	[0.235]	[0.129]	[0.341]	[0.266]
HSR×*lndistance*	0.001	-0.178*	0.006	-0.040	-0.011	0.007
	(0.038)	(0.093)	(0.052)	(0.031)	(0.056)	(0.042)
	[0.038]	[0.090]	[0.047]	[0.027]	[0.068]	[0.053]
Observations	1,896	1,896	1,896	1,896	1,896	1,896
R-squared	0.703	0.681	0.582	0.594	0.370	0.360
Predetermined-year Control	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓

Notes: Table 3 reports the baseline regression results for different kinds of protest events based on contents. Column (1) reports the baseline results. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR connectivity and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. Robust standard errors in parentheses are clustered at the prefecture levels. Standard errors in brackets are Conley standard errors robust for spatial correlation, assuming a cutoff window of 500 km and a serial correlation of 6 years. ***p< 0.01, **p <0.05, *p < 0.1.

5.3 Alternative Data Source

To ensure the robustness of our results and mitigate potential biases from social media, we validate our findings using the CLB dataset, which is based on actual occurrences of worker-led protests. Given that A-to-NA migrant workers are the primary contributors to our observed effects, our results should consistently apply to worker protests in the CLB, as opposed to large-scale anti-state protests, such as those captured in the GDELT dataset. Table 4 illustrates HSR's impact on strike frequency, showing results consistent with our baseline: HSR does not increase nor reduce strikes per se, but it reveals pronounced spatial heterogeneity. As shown in column (2), HSR opening leads to an increase of 3 strikes in central regions, with a reduction in peripheral areas. Pre-trend tests are shown in Figure C4. With insignificant change prior to HSR opening, strikes soar in central regions (in Panel A and B) but dwindle with the geographical distance (in Panel C and D). Thus, there is no clear pre-trends, and the effect persists for long.

To further explore the role of A-to-NA migrants, we examine the industries in which strikes take place. We classify the industries into two groups: pro-migrant (construction, manufacturing, services, and mining) and other industries (government agencies and the education sector). A-to-NA migrants, typically low skilled but with strong laborforce, predominantly work in the pro-migrant ones. If they are driving our findings, we would expect HSR to have a significant effect on strikes in pro-migrant ones, but little effect on others. Columns (3) and (4) of Table 4 confirm this hypothesis: HSR openings increase strikes in pro-migrant industries in central regions while reducing them in remote areas, but there is no significant effect on strikes in other industries. Overall, labor strikes in pro-migrant industries respond to HSR openings in a manner similar to protests in the baseline, highlighting the pivotal role of A-to-NA migrant workers.

5.4 Robustness Checks

In this section, we briefly discuss a few important robustness checks, with a more detailed analysis elucidated in Appendix D.

TABLE 4 Baseline Results: CLB dataset

	(1) Strike	(2) Strike	(3) Strike in pro-migrants industries	(4) Strike in other industries
HSR	0.108	2.999***	2.596***	0.192
	(0.368)	(0.940)	(0.930)	(0.156)
	[0.336]	[0.891]	[0.814]	[0.168]
HSR×*lndistance*		-0.646***	-0.554***	-0.045
		(0.192)	(0.190)	(0.032)
		[0.184]	[0.168]	[0.035]
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	2,456	2,438	2,438	2,438
R-squared	0.801	0.814	0.806	0.362

Notes: Table 4 reports the baseline regression results using CLB dataset during 2010-2018. The dependent variable is the total number of strikes at the prefectural level in columns (1)-(2), the number of strikes in industries that are more attractive to migrant workers, such as construction, manufacturing, services, and mining in column (3), and the number of strikes in industries that are less likely to accept migrant workers, such as party and government agencies, and the education sector in column (4). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. The predetermined-year control variables include the interaction terms between the year dummy variables and the proportions of the primary sector, secondary sector, per capita GDP, population, and urban area in 2010. Standard errors in brackets are Conley standard errors robust for spatial correlation, assuming a cutoff window of 500 km and a serial correlation of 6 years. ***p< 0.01, **p <0.05, *p < 0.1.

Placebo test. We conduct a series of placebo tests to validate our findings. Specifically, we replace distance to the provincial capital with alternative measures (e.g., distance to Beijing, major economic centers, or non-capital cities, See Table D1), employ historical administrative and trade centers (Table D2), and randomly permute HSR opening years (Figure D1). Across all specifications, the results remain insignificant, reinforcing that the spatial pattern of protest emerges only relative to provincial capitals.

Long-difference, market-access, and recentered-IV. To address potential biases from transitory shocks or serial correlation (Griliches and Hausman, 1986), we construct a cross-sectional dataset for 2010–2016 and examine how changes in HSR access relate to changes in protest events (columns (1)–(2) of Table D3). To capture the quality of connectivity, following Borusyak and Hull (2023), we employ a continuous measure of HSR exposure based on market access (MA) growth (columns (3)–(4)). To address endogeneity concerns, we further implement a recentered-IV approach (columns (5)–(6)), which isolates exogenous variation in MA growth. We validate this instrument by showing that re-centered MA growth is uncorrelated with expected MA growth (Figure D2). Across these specifications, our conclusions remain robust, even when accounting for the anticipated effects of HSR construction.

Excluding confounding factors. We consider the potential influence of officials' personal characteristics in Table D4. In Table D5, We exclude the confounding effects of alternative transportation modes, and in Table D6, we control for the Internet penetration rate to alleviate potential measurement errors. All results consistently indicate the robustness of our baseline findings[20].

Other Robustness Checks. A critical assumption underlying our analysis is the overlap between the economic and political centers, with provincial capitals possessing strong economic centripetal power. While this pattern holds for most provincial capitals, we test the sensitivity of our results to some exceptions in Table D7. These tests account for various scenarios where centripetal power

20. We further control for a wide range of potential confounders—including whether local officials are natives, local fiscal conditions, cultural norms and social organizations, natural geographical factors, and weather conditions—and find our results remain robust. For brevity, these additional results are not reported here.

may be weaker, and the results remain consistent with our baseline. Combined with Table D1, the results endorse this assumption and support our focus on within-province outcomes. Additionally, we examine the robustness of our findings to different model specifications. The results in Figure D3, Figure D4, Table D8, and Table D9 show that our conclusion is robust to alternative time windows, clustering methods, dependent variables, and distance measurements.

6 Possible Channels and Alternative Explanations

6.1 Monitoring or Deterrence Effect?

In this subsection, we find little evidence that improved government monitoring drives our finding. Instead, the evidence is more consistent with the deterrence effect, where expanded state presence raises the perceived costs of protest participation.

To begin with, we test the influence of HSR on investigations into local officials. Since 2012, CPC has conducted a sweeping anti-corruption campaign, sending inspection teams to every province (usually stationed at the provincial center), followed by further investigation teams to local prefectures. If the monitoring effect holds, local officials in areas with HSR should be more exposed to investigation compared to those without HSR. We manually collect the data from local governments' websites, cross-verified by social media, to obtain a large dataset on whether the prefectural officials were investigated (*Investigate*), and the weighted average of their rankings (*Weighted Investigate*).[21] However, columns (1) and (2) of Table E1 show little influence by HSR, whether in central regions or remote areas. We further include interaction terms to test their influence on protests. Columns (3) and (4) also show negligible effects. In general, HSR opening can hardly alleviate corruption, so the monitoring effect (Hypothesis 1) fails to explain our findings.

We then provide some suggestive evidence for the deterrence effect proposed in Hypothesis 2

21. The dataset covers all investigated officials ranking above the county level. Note that in the context of China, only those investigated to have violated party discipline or law will be made public, so "investigation" here is equivalent to prosecution and legal punishment. We denote a prefecture "investigated" if its main officials were investigated. The weighted average is calculated by averaging the ranking of the investigated officials in the prefecture. For prefecture p, $WI_p = \sum_{i \in I} i \frac{\#(\text{investigated officials with rank i in prefecture p})}{\#(\text{investigated officials in prefecture p})}$. The rank i is the official administrative range of the Chinese government, ranging from 1 to 10 ($I = \{1, 2, ..., 10\}$), representing from vice township level to country level. The ranks of prefecture officials lie primarily between 5 (vice-prefecture level) and 7 (vice-province level).

We employ the ISCIU dataset to examine how HSR reshapes their political expressions, focusing on two prominent forms: writing to public media or engaging in protests. The deterrence effect predicts a decline in these forms of expression in remote areas after HSR opening. Table 5 confirms the hypothesis: people in remote areas are less inclined to express their views after HSR opening, implying that stronger state presence deters political expression.

TABLE 5 Suggestive evidence for the deterrence effect.

	(1)	(2)	(3)	(4)
	Expressing political opinions through			
Dependent variable	Writing articles to the media		Protests	
HSR	0.104	0.223***	0.079	0.231***
	(0.077)	(0.075)	(0.073)	(0.072)
HSR×*lndistance*	-0.028***	-0.036***	-0.017**	-0.032***
	(0.009)	(0.009)	(0.008)	(0.008)
Individual Controls		✓		✓
Province FE	✓	✓	✓	✓
Observations	2,339	2,339	2,339	2,339
R-squared	0.028	0.029	0.081	0.103
Outcome mean	1.745	1.745	1.696	1.696

Notes: Table 5 reports the suggestive evidence for the deterrence effect. ***p< 0.01, **p <0.05, *p < 0.1. The data is primarily sourced from the Ideology Survey of Chinese Internet Users (ISCIU) dataset. Robust standard errors are clustered at the prefecture levels. The dependent variables in columns (1)-(2) represent the frequency at which individuals choose to express their political opinions through writing and submitting articles to the media. The dependent variables in columns (3)-(4) represent the frequency at which individuals express their political opinions through participating in protests and demonstrations. The variables range from 1 to 4, with higher values indicating a higher frequency. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital.

6.2 Agglomeration or Spillover Effect?

While state infrastructure reduces protest in peripheral regions, it simultaneously gives rise to unrest in urban centers. In this subsection, we provide evidence that the increase in protests following the HSR in provincial capitals is primarily driven by the inflow of A-to-NA migrants. First, we confirm HSR's positive effect on both population and economic agglomeration in central regions, supporting the first channel. We then falsify the second channel by showing that HSR opening does not decentralize resources to remote areas. Second, we find that HSR increases

protests related to migrants, and they contribute mostly to our observed effects. Moreover, the affected migrants are low-skilled laborforce from rural areas, who may face severe discrimination. In sum, we highlight the role of A-to-NA migrants in driving our main findings.

TABLE 6 The heterogeneous impact of HSR openings on agglomeration

Dependent variable	(1) *Float ratio*	(2) *Population agglomeration*	(3) *Economic agglomeration*	(4) *GDP Per capita*
HSR	0.065***	0.024*	0.083**	7.813***
	(0.021)	(0.013)	(0.032)	(1.563)
	[0.016]	[0.013]	[0.026]	[1.966]
HSR×*lndistance*	-0.015***	-0.005*	-0.018**	-1.678***
	(0.004)	(0.003)	(0.007)	(0.316)
	[0.003]	[0.003]	[0.005]	[0.385]
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,753	1,913	1,913	1,940
R-squared	0.826	0.999	0.990	0.923
Outcome mean	-0.044	5.730	3.389	44.678

Notes: Table 6 reports the heterogeneous impact of HSR openings on agglomeration. The dependent variables in columns (1) is the float ratio (the proportion of migrants). The dependent variables in columns (2)-(3) are the indices of population agglomeration, economic agglomeration, which are measured by the ratios of year-end population, employment in the prefecture to their respective national totals, as well as the ratios of the prefecture's administrative area to the national total. In column (4), the dependent variable is per capita GDP (in thousands). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. Robust standard errors in parentheses are clustered at the prefecture levels. Standard errors in brackets are Conley standard errors robust for spatial correlation, assuming a cut-off window of 500 km and a serial correlation of 6 years. ***p< 0.01, **p <0.05, *p < 0.1.

6.2.1 Economic and Population Agglomeration

We first examine how HSR affects population flows. Column (1) in Table 6 reports results using float ratio (the proportion of migrants) as the dependent variable. We find that HSR opening significantly increases the float ration in the provincial capital, which implies that more migrants come to central regions. This agglomeration effect diminishes with increasing distance to provincial capitals, and in some remote regions, the float ratio even declines. Figure E1 plots the dynamic effects, which shows no pre-trends and indicates that the proportion of migrants in provincial capitals increases steadily after HSR opening. The results suggest that HSR opening induces more

floating population to central regions.

We then construct measures to test whether HSR openings foster agglomeration. We measure population/economic agglomeration by the ratio of the prefecture's population/employment fraction to its administrative area fraction (the fraction is defined with regard to the national value). The results are presented in columns (2) to (4) in Table 6, showing that HSR opening leads to higher population, employment, and GDP in central regions, thus reshaping the economic geography of the whole province. In sum, the agglomeration effect contributes to our findings.

To testify the spillover effect, we compare the average wages and housing prices before and after HSR opening. According to Hypothesis 4, HSR should lead to lower wages or higher housing prices in central regions but the opposite in peripheral areas. However, columns (4)-(5) in Table E2 show ambiguous results, with no significant changes in either wages or housing prices. Furthermore, HSR does not bring about significant changes in industrial structure, so the observed changes cannot be driven by composition.[22] Taken together, the results suggest that our findings are not due to improved economic conditions of peripheral regions proposed by Hypothesis 4. Instead, increasing job opportunities in central areas induced by HSR could attract more migrants, leading to ever-rising agglomeration, which validates Hypothesis 3.

6.2.2 The role of A-to-NA migrants

We then explore the role of A-to-NA migrants in driving our findings. We present both direct and indirect evidence to support this channel. First, we document that protests involving A-to-NA migrants rather than non-migrant protest drive our main results. Second, we use micro-survey data to show that HSR-related migration flows predominantly consist of low-skilled workers.

First, we employ text analysis methods to classify the content of all protest events in the CASM dataset and extract those involving A-to-NA migrants. In Table 7, columns (1)-(2) define migrant-related protests as those involving migrant workers, while columns (3)-(4) further restrict the definition to protests related to both migrant workers and *hukou* issues, which is the most severe

22. An, Qin, Wu, and You (2024) show the relaxation of *hukou* system reduces wages on migrants instead of urban residents, which again corroborates our findings.

problem that they encounter (An et al., 2024). As shown in Panel A of Table 7, protests involving A-to-NA migrants exhibit a similar trend to the baseline, increasing in central regions but decreasing in peripheral areas. In contrast, when we exclude these migrant-related protests and use non-migrant protests as dependent variables in Panel B, the results turn insignificant for both terms. This suggests that migrant protests are the primary driving force of our findings.

TABLE 7 The impact of HSR opening on protest involving migrant workers.

Definition	*Migration worker-related keywords*		*Migration worker-Hukou-related keywords*	
	(1)	(2)	(3)	(4)
	Panel A: Migration-related Protest			
HSR	1.035***	0.840*	1.026***	0.830*
	(0.369)	(0.427)	(0.369)	(0.428)
HSR×*lndistance*	-0.222***	-0.178**	-0.221***	-0.177*
	(0.079)	(0.090)	(0.079)	(0.090)
	Panel B: Non-Migration-related Protest			
HSR	-0.256	-0.428	-0.251	-0.420
	(0.381)	(0.382)	(0.384)	(0.385)
HSR×*lndistance*	0.047	0.080	0.046	0.079
	(0.080)	(0.081)	(0.081)	(0.081)
Controls		✓		✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,943	1,896	1,943	1,896

Notes: Table 7 reports the impact of HSR opening on protest involving migrant workers. In Panel A, the dependent variable is the standardized number of protests related to A-to-NA migrant workers at the prefecture-year level. In Panel B, the dependent variable is the standardized number of protests unrelated to A-to-NA migrant workers. In columns (1)-(2), migrant-related protests are defined as those involving migrant workers. In columns (3)-(4), migrant-related protests further include those related to *hukou* issues. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. Robust standard errors are clustered at the prefecture levels. ***p< 0.01, **p <0.05, *p < 0.1.

To further explore this channel, we examine the main features of these A-to-NA migrants, using data on 900,000 migrants from 2011 to 2016 from CMDS. We employ the number of HSR openings in other prefectures within the same province to proxy for the extension of HSR network from the provincial capital, and analyze its relationship with the migrant features in provincial capitals. Figure E2 reveals that, as more prefectures become connected to HSR within the province, the average education level of the migrants decreases, while the proportion with rural *hukou* increases. This suggests that HSR mainly drives the inflow of low-skilled A-to-NA migrants from peripheral

areas. We find no significant changes in terms of gender, age, or ethnicity.

6.3 Testing Alternative Explanations

HSR-related protests. We consider protests launched directly related to HSR as the construction may trigger unrest due to improper route choices or insufficient compensation. These protests should have peaked shortly after HSR opening and declined over time. However, this is inconsistent with our findings reported in Figure 3 that protests in central regions increased gradually over time. We further test this explanation by two approaches. First, we classify those protest related to HSR construction, examining its variation with HSR opening, and exclude them to show the effect on non-HSR protests. Table E3 shows that the effect is insignificant for HSR-related protests, but highly significant for the remaining non-HSR protests. Second, we code HSR access as non-absorbing to assess its immediate impact. Column (7) shows a insignificant result, which again shows that instant HSR-related protests can hardly drive our baseline findings.

Commuting for protest. Another possible explanation is that HSR may serve as a "coordination devise", helping protesters commute to the capital for petitioning to the provincial government. If so, the effect should be most evident for high-ranking issues where provincial bureaucrats are especially important. However, the results in Table 3 suggest that this is unlikely, as the protests we observe are mostly economic issues, rather than ethnic conflicts or veteran problems that may be of higher political importance. We also test the case by identifying protests relating to upward petitioning (the keyword "*shangfang*"). Figure E3 shows that these protests are hardly influenced by HSR opening. This excludes the possibility that protesters merely commute to central regions for petitioning.

Changes in public good provision. Another explanation is that local governments are provided with more transfers to support HSR construction, thus delivering better public goods. To rule out this concern, we test the influence of HSR opening on the provision of public goods in Table E4, including the per capita number of hospital beds, Internet access, parks (area), public books, buses,

the proportion of education expenditure or science and technology expenditure in fiscal revenue. We find no evidence that HSR opening could increase the provision of these public goods.

7 Conclusion and Discussion

Transportation projects have long stood as visible manifestations of state power, from canals in Mesopotamia to highway networks in the United States and the recent HSR expansion in China. While conventional wisdom treats them as instruments of state presence and control, this paper demonstrates an *authoritarian infrastructure paradox*: while infrastructure can extend state presence and deter dissent in peripheral regions, it can simultaneously relocate protest to politically sensitive centers, posing risks to regime stability. Drawing on detailed protest data from China, we provide robust empirical evidence consistent with this paradox.

This study advances research on the politics of infrastructure and authoritarian governance in two ways. First, we move beyond the question of whether infrastructure simply increases or decreases protest, and instead show that it fundamentally reshapes the *political geography* of protest. Second, by combining a formal model with causal evidence from China's HSR program, we demonstrate how protests migrate from the periphery to the center, extending prior works on spatial politics and state presence (Nall, 2015; Schuler and Truong, 2020).

Our results can readily be generalized to other countries or different infrastructure projects, though caution is necessary when extending these results due to the assumption that economic and political centers overlap. In countries or regions where they are separated, political power and economic forces function independently, and even identical policy shock could exhibit distinct political implications and spatial patterns. We expect more research to explore the interplay between state power and economic forces in shaping the spatial distribution of political economy.

References

Acemoglu, D., C. García-Jimeno, and J. A. Robinson (2015). State capacity and economic development: A network approach. *American Economic Review 105*(8), 2364–2409.

Akbulut-Yuksel, M., D. Okoye, and B. Turan (2024). Expressway to votes: Infrastructure projects and voter persuasion. *Economic Journal 134*(657), 48–94.

Amos, P., P. Amos, D. Bullock, and J. Sondhi (2010). *High-Speed Rail: The Fast Track to Economic Development?*. World Bank Washington, DC, USA.

An, L., Y. Qin, J. Wu, and W. You (2024). The local labor market effect of relaxing internal migration restrictions: evidence from china. *Journal of Labor Economics 42*(1), 000–000.

Anderson, C. J. and Y. V. Tverdova (2003). Corruption, political allegiances, and attitudes toward government in contemporary democracies. *American Journal of Political Science 47*(1), 91–109.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang (2017). Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics 99*(3), 435–448.

Bautista, M. A., F. González, L. R. Martínez, P. Muñoz, and M. Prem (2023). The geography of repression and opposition to autocracy. *American Journal of Political Science 67*(1), 101–118.

Beissinger, M. R. (2022). *The revolutionary city: Urbanization and the global transformation of rebellion*. Princeton University Press.

Bernard, A. B., A. Moxnes, and Y. U. Saito (2019). Production networks, geography, and firm performance. *Journal of Political Economy 127*(2), 639–688.

Blair, C. W. (2025). Border fortification and legibility: Evidence from afghanistan. *American Journal of Political Science 69*(4), 1559–1580.

Borusyak, K. and P. Hull (2023). Non-random exposure to exogenous shocks. *Econometrica 91*(6), 2155–2185.

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

Cai, Y. (2008, July). Power Structure and Regime Resilience: Contentious Politics in China. *British Journal of Political Science 38*(3), 411–432.

Cai, Y. (2010, February). *Collective Resistance in China: Why Popular Protests Succeed or Fail*. Stanford University Press.

Calamunci, F. and J. Lonsky (2025). The road to crime: An unintended consequence of the interstate highway system. *Economic Journal 135*(667), 748–772.

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

Campante, F. R., D. Chor, and B. Li (2023). The political economy consequences of china's export slowdown. *Journal of the European Economic Association 21*(5), 1721–1771.

Campante, F. R., Q.-A. Do, and B. Guimaraes (2019). Capital cities, conflict, and misgovernance. *American Economic Journal: Applied Economics 11*(3), 298–337.

Cantoni, D., A. Kao, D. Y. Yang, and N. Yuchtman (2024). Protests. *Annual Review of Economics 16*, 519–543.

Cao, Y. and S. Chen (2022). Rebel on the canal: Disrupted trade access and social conflict in china, 1650–1911. *American Economic Review 112*(5), 1555–1590.

Chang, C. and Y. Wang (2024). The reach of the state. *Comparative Political Studies 57*(8), 1243–1275.

Charnoz, P., C. Lelarge, and C. Trevien (2018). Communication costs and the internal organisation of multi-plant businesses: evidence from the impact of the french high-speed rail. *Economic Journal 128*(610), 949–994.

Chen, T., J. K.-s. Kung, and C. Ma (2020). Long live keju! the persistent effects of china's civil examination system. *Economic Journal 130*(631), 2030–2064.

Chiovelli, G., L. Fergusson, L. R. Martínez, J. D. Torres, and F. V. Caicedo (2024). Bourbon reforms and state capacity in the spanish empire. *Working Paper*.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Crost, B., J. Felter, and P. Johnston (2014). Aid under fire: Development projects and civil conflict. *American Economic Review 104*(6), 1833–1856.

Davis, S. J., M. Qian, and W. Zeng (2025). A comprehensive gis database for china's surface transport network with implications for transport and socioeconomics research. Technical report, National Bureau of Economic Research.

De Chaisemartin, C. and X. d'Haultfoeuille (2024, 02). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, 1–45.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review 108*(4-5), 899–934.

Dong, X. (2018). High-speed railway and urban sectoral employment in china. *Transportation Research Part A: Policy and Practice 116*, 603–621.

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

Duranton, G. and D. Puga (2015). Urban land use. In *Handbook of regional and urban economics*, Volume 5, pp. 467–560. Elsevier.

Elfstrom, M. (2021, January). *Workers and Change in China: Resistance, Repression, Responsiveness* (1 ed.). Cambridge University Press.

Enikolopov, R., A. Makarin, and M. Petrova (2020). Social media and protest participation: Evidence from russia. *Econometrica 88*(4), 1479–1514.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies 81*(3), 1046–1070.

Fang, H., M. Li, L. Wang, and Y. Yang (2025). High-speed rail and china's electric vehicle adoption miracle. Technical report, Penn Institute for Economic Research, Department of Economics, University of ....

Fu, D. (2017, November). *Mobilizing without the Masses: Control and Contention in China*. Cambridge University Press.

Glaeser, E. L., J. D. Gottlieb, and K. Tobio (2012). Housing booms and city centers. *American Economic Review 102*(3), 127–133.

González, F., J. Miquel-Florensa, M. Prem, and S. Straub (2025). The dark side of infrastructure: Roads, repression and land in authoritarian paraguay. *Economic Journal 135*(666), 653–669.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology 83*(6), 1420–1443.

Griliches, Z. and J. A. Hausman (1986). Errors in variables in panel data. *Journal of Econometrics 31*(1), 93–118.

Gumpert, A., H. Steimer, and M. Antoni (2022). Firm organization with multiple establishments. *Quarterly Journal of Economics 137*(2), 1091–1138.

Gurr, T. R. (1970). *Why Men Rebel*. Princeton University Press.

Huntington, S. P. (2006). *Political order in changing societies*. Yale university press.

Kalyvas, S. N. (2006). *The logic of violence in civil war*. Cambridge University Press.

Le Bon, G. (1897). *The crowd: A study of the popular mind*. Courier Corporation.

Lee, C. K. (2007). *Against the law: labor protests in China's rustbelt and sunbelt*. Univ of California Press.

Lee, C. K. and Y. Zhang (2013). The power of instability: Unraveling the microfoundations of bargained authoritarianism in china. *American Journal of Sociology 118*(6), 1475–1508.

Li, H. and L.-A. Zhou (2005). Political turnover and economic performance: the incentive role of personnel control in china. *Journal of Public Economics 89*(9-10), 1743–1762.

Li, Y. (2019). *Playing by the Informal Rules: Why the Chinese Regime Remains Stable despite Rising Protests*. Cambridge University Press.

Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from china's high speed railway system. *Journal of Urban Economics 98*, 98–123.

Liu, L., X. Cao, S. Li, and N. Jie (2024). A 31-year (1990–2020) global gridded population dataset generated by cluster analysis and statistical learning. *Scientific Data 11*(1), 124.

Ma, D. and L. Wang (2015). Chinese netizens' political ideology and their sources: an empirical analysis. *Chinese Journal of Sociology 35*(5), 142–167.

Ma, X. (2022). *Localized bargaining: The political economy of China's high-Speed Railway program*. Oxford University Press.

Manacorda, M. and A. Tesei (2020). Liberation technology: Mobile phones and political mobilization in africa. *Econometrica 88*(2), 533–567.

Mann, M. (1984). The autonomous power of the state: its origins, mechanisms and results. *European Journal of Sociology/Archives européennes de sociologie 25*(2), 185–213.

Mastrorocco, N. and E. Teso (2023). State capacity as an organizational problem. evidence from the growth of the us state over 100 years. *Working Paper*.

Mattingly, D. C. (2016). Elite capture: How decentralization and informal institutions weaken property rights in china. *World Politics 68*(3), 383–412.

Morris, S. and H. S. Shin (2000). Rethinking multiple equilibria in macroeconomic modeling. *NBER macroeconomics Annual 15*, 139–161.

Müller-Crepon, C., P. Hunziker, and L.-E. Cederman (2021). Roads to rule, roads to rebel: Relational state capacity and conflict in africa. *Journal of Conflict Resolution 65*(2-3), 563–590.

Nall, C. (2015). The political consequences of spatial policies: How interstate highways facilitated geographic polarization. *Journal of Politics 77*(2), 394–406.

O'Brien, K. J. and L. Li (2006). *Rightful Resistance in Rural China*. Cambridge Studies in Contentious Politics. Cambridge: Cambridge University Press.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in indonesia. *Journal of Political Economy 115*(2), 200–249.

Pan, Y., N. Shroff, and P. Zhang (2023). The dark side of audit market competition. *Journal of Accounting and Economics 75*(1), 101520.

Passarelli, F. and G. Tabellini (2017). Emotions and political unrest. *Journal of Political Economy 125*(3), 903–946.

Pengl, Y. I., C. Müller-Crepon, R. Valli, L.-E. Cederman, and L. Girardin (2025). The train wrecks of modernization: Railway construction and separatist mobilization in europe. *American Political Science Review*, 1–18.

Polanyi, K. et al. (1944). The great transformation: The political and economic origins of our time.

Pruett, L. (2024). Resisting the blood tax: Coercive capacity, railroads, and draft evasion in colonial west africa. *Journal of Politics 86*(4), 1207–1220.

Qin, B., D. Strömberg, and Y. Wu (2024). Social media and collective action in china. *Econometrica 92*(6), 1993–2026.

Qin, Y. (2017). No county left behind? the distributional impact of high-speed rail upgrades in china. *Journal of Economic Geography 17*(3), 489–520.

Rho, S. (2023). *Atomized Incorporation: Chinese Workers and the Aftermath of China's Rise*. Cambridge University Press.

Rogowski, J. C., J. Gerring, M. Maguire, and L. Cojocaru (2022). Public infrastructure and economic development: evidence from postal systems. *American Journal of Political Science 66*(4), 885–901.

Schuler, P. and M. Truong (2020). Connected countryside: the inhibiting effect of social media on rural social movements. *Comparative Politics 52*(4), 647–669.

Seligson, M. A. (2002). The impact of corruption on regime legitimacy: A comparative study of four latin american countries. *Journal of Politics 64*(2), 408–433.

Sexton, R. and C. Zürcher (2024). Aid, attitudes, and insurgency: Evidence from development projects in northern afghanistan. *American Journal of Political Science 68*(3), 1168–1182.

Shue, V. (1990). *The reach of the state: sketches of the Chinese body politic*. Stanford University Press.

Steinhardt, H. C. and F. Wu (2016, January). In the Name of the Public: Environmental Protest and the Changing Landscape of Popular Contention in China. *China Journal 75*, 61–82.

Steinhardt, H. C. and K. Yang (2025). Value-driven contention in china: Forms, tactics and state responses. *China Quarterly 261*, 39–56.

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

Tilly, C. (2003). *The politics of collective violence*. Cambridge University Press.

Tombe, T. and X. Zhu (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review 109*(5), 1843–1872.

Traugott, M. (2010). *The insurgent barricade*. University of California Press.

Voigtländer, N. and H.-J. Voth (2025). Highway to hitler. *American Economic Journal: Applied Economics*, forthcoming.

Wallace, J. L. (2014). *Cities and stability: Urbanization, redistribution, and regime survival in China*. Oxford University Press.

Wang, Z., Q. Zhang, and L.-A. Zhou (2020). Career incentives of city leaders and urban spatial expansion in china. *Review of Economics and Statistics 102*(5), 897–911.

Weber, M. (2016). Economy and society. In *Democracy: A Reader*, pp. 247–251. Columbia University Press.

Xu, C. (2011). The fundamental institutions of china's reforms and development. *Journal of Economic Literature 49*(4), 1076–1151.

Xu, X. (2021). To repress or to co-opt? authoritarian control in the age of digital surveillance. *American Journal of Political Science 65*(2), 309–325.

Xu, Y. and Y. Yao (2015). Informal institutions, collective action, and public investment in rural china. *American Political Science Review 109*(2), 371–391.

Ying, L. (2021). How state presence leads to civil conflict. *Journal of Conflict Resolution 65*(2-3), 506–533.

Zhang, H. and J. Pan (2019). Casm: A deep-learning approach for identifying collective action events with text and image data from social media. *Sociological Methodology 49*(1), 1–57.

Zheng, S. and M. E. Kahn (2013). China's bullet trains facilitate market integration and mitigate the cost of megacity growth. *Proceedings of the National Academy of Sciences 110*(14), 1248–1253.

Appendix for "Where State Presence Backfires: How State Infrastructure Reshapes Protest Geography"

(For Online Publication)

Appendix A Theoretical Framework: Formal Analysis **A1**

A.1 Settings A1

A.2 A Simple Model of Collective Resistance A2

A.3 Agglomeration Effect or Spillover Effect A3

A.4 Monitoring Effect or Deterrence Effect A5

A.5 Proof for existence and uniqueness A6

Appendix B More on Background and Data **A8**

Appendix C More on Empirical Results **A13**

Appendix D More on Robustness Checks **A20**

D.1 Placebo Test A20

D.2 Recenter IV Approach, Long-difference Setting and Market Access Exposure A23

D.3 More Controls A26

D.4 Sub-sample Tests: Possible Outliers A30

D.5 Alternative Model Specifications A32

Appendix E More on Channels **A35**

Appendix A Theoretical Framework: Formal Analysis

In this section, we build a spatial model to explore how HSR affects protest incidence, following the framework of political riots (Passarelli and Tabellini, 2017).

A.1 Settings

Consider an economy consisting of one central city and a large number of small cities around (with linear form, as in Wang, Zhang, and Zhou (2020)). In our setting, the center is the provincial capital, while the small cities are prefectures surrounding the capital. Denote each city (or prefecture) by i, with d_i and N_i representing its distance to the capital, its population, respectively. Let $N = \sum_{i=1}^{n} N_i$ be the total population and D denoting the distance of the farthest area. In other words, $0 \leq d_i \leq D, \forall i \in \mathcal{N} = \{1, ..., n\}$, where $d_i = 0$ for the capital and $d_i = D$ for the most remote city. Suppose the labor participation rate is $\alpha \in (0, 1)$, then αN_i is the average population of labor in city i.[1]

Since the circular model is isomorphic to a linear one,[2] we simplify the model to a line, with the center located at the origin "0" and other cities located at $d_i \leq D$. For a certain individual j living in city i while working in city k, his[3] utility is a continuous, increasing function of consumption (or surplus) $u_{ik}^j = u(c_{ik}^j)$. On average, $\mathbb{E}(c_{ik}^j) = c_{ik} = w_k - R_k - n_{ik}T_{ik}(d_{ik})$, where w_k, R_k, n_{ik}, T_{ik} represents the average wage in city k, land rent in k, the average mobility frequency between k and i, and the mobility cost. $d_{ik} = |d_i - d_k|$ is the geographical distance.

Concerning the strong economic centripetal force of the provincial capital and our focus on HSR, we simplify the model by assuming people can only make a binary work choice between his hometown and the capital. After migrating to the capital, he gets $c_{i0} = w_0 - R_0 - n_{i0}T_{i0}(|d_i|)$. Alternatively, he can work locally and gain $c_{ii} = w_i - R_i$. We hold constant the mobility frequency n_{i0} and focus on the mobility cost $T_{i0}(|d_i|)$ in our subsequent analyses. We assume $T_{i0}(\cdot)$ to be an increasing function without specifying a functional form.

We then provide a basic setting for protests. Each individual j unilaterally decides whether or not to participate in protests faced with aggrievement a_{ij}. While aggrievement has many sources, we focus on the gap between expected entitlements $\bar{u}$ and delivered utility u_{ij}. This gap can be purely emotional Passarelli and Tabellini (2017), signifying the individual's "perceived" discrimination. Individuals are more aggrieved when faced with greater utility gap, so $a_{ij} = a(\bar{u} - u_{ij})$ is decreasing in u_{ij} when $u_{ij} \leq \bar{u}$ (while holding constant when the bar is reached). We assume

1. There are $(1-\alpha)N_i$ people out of laborforce, like children and the elderly. We may also consider $\alpha = \mathbb{E}(\alpha_i)$, where α_i is the labor participation rate of each city i.

2. The isomorphism is constructed by taking an equivalence class based on distance. Let $\mathcal{C}$ be the original disk, then $\mathcal{L} = \mathcal{C}/\mathcal{D}$, where $\mathcal{D} = \{i \in \mathcal{N} : d_i = d\}$. Hence, $\mathcal{C}$ and $\mathcal{L}$ are isomorphic based on the quotient mapping $\pi : \mathcal{C} \to \mathcal{L}$.

3. We use "he" for residents and "she" for officials for concision, without special reference.

$\bar{u}$ to be large enough so that a_{ij} is decreasing in u_{ij}. Denote with p_i the average participation rate in protests for city i. In the next subsection we derive the equilibrium participation rate that can be expressed as a function of common aggrievement, then we connect it with HSR opening and discuss the potential channels for our empirical findings.

A.2 A Simple Model of Collective Resistance

Individuals unilaterally decide whether to participate in resistance, trading off the cost and benefit of participation. Based on extant literature (Gurr, 1970; Granovetter, 1978), we model the benefit as emotional: it is the psychological reward of joining other residents in a public display of the aggrievement. Let a_i denote the common aggrievement among residents in city i. Let λ_i denote the number of potential protesters, which is the residents (total population plus/minus the migrants) living in city i and adjacent cities. We assume the mobilization range to be $r \in (0, D)$ so that a protest occurring in city i can at most mobilize residents living in $j \in \{j : |d_i - d_j| \leq r\}$. We assume there is a complementarity: the benefit of participation grows proportionately with the mass of other participants $p_i\lambda_i$. This assumption is plausible on several grounds. The psychological benefit of a public display of anger is likely to be stronger if more people join; alternatively, participation could proxy for the probability that a critical threshold of participators is reached such that the crowd overwhelms the police (Morris and Shin, 2000). Equivalently, on the cost side, the probability of being arrested is smaller in a large crowd (Le Bon, 1897).

On the other hand, joining a resistance entails costs, in terms of time, or risk of being arrested or injured. We model these costs as the sum of three components, $\mu + \mu_i + \epsilon_{ij}$. The parameter $\mu > 0$ is known ex-ante, common to all individuals and cities, reflecting external conditions like the time cost. $\mu_i > 0$ is common to all individuals in city i, reflecting the probability of repression (perceived deterrence). The term ϵ_{ij} is a random variable that captures idiosyncratic components of the cost, following a distribution F_i. This distribution is common knowledge, continuous, and has density f_i, with support lying on both sides of 0.

Combining these assumptions, individual j in city i chooses to join the crowd in protests if benefits are larger than costs: $p_i\lambda_i a_i - \mu - \mu_i - \epsilon_{ij} \geq 0$. The fraction of individuals in city i who participate in the protest is thus given by:

$$p_i = Pr(\epsilon_{ij} \leq p_i\lambda_i a_i - \mu - \mu_i) = F_i(p_i\lambda_i a_i - \mu - \mu_i). \tag{A1}$$

Hence, the equilibrium participation rate p_i^* is a fixed point of Equation A1. We focus on the interior solution whose existence and uniqueness is proved in subsection A.5.

We can express $p_i^* = H_i(a_i)$, with other parameters subsumed in the $H_i(\cdot)$ function. The com-

parative statics of Equation A1 implies:

$$H_i^{'} = \frac{\partial H_i(\cdot)}{\partial a_i} = \frac{\partial p_i^*}{\partial a_i} = \frac{\lambda_i p_i f_i}{1 - a_i \lambda_i f_i} \tag{A2}$$

By Assumption A2, $1 - a_i\lambda_i f_i > 0$, i.e., $H_i^{'} = \frac{\partial p_i^*}{\partial a_i} > 0$. Hence, H_i is strictly increasing in a_i, which means the protest rate is higher with stronger aggrievement. Hence, people protest less when faced with less corruption or receiving a higher surplus.

Equation A2 implies that $\frac{\partial^2 p_i^*}{\partial a_i \partial \lambda_i} = \frac{a_i f_i}{(1-a_i\lambda_i f_i)^2} > 0$. Thus, participation is more sensitive to aggrievement with larger λ_i, i.e., there is a greater number of potential protesters in the group. With stronger economic agglomeration, an individual expects more people will be involved and would be more willing to join.

Similarly, we have

$$\frac{\partial H_i(\cdot)}{\partial \lambda_i} = \frac{\partial p_i^*}{\partial \lambda_i} = \frac{a_i p_i f_i}{1 - a_i \lambda_i f_i}. \tag{A3}$$

It is evident that $\frac{\partial p_i^*}{\partial \lambda_i} > 0$. This shows more potential protesters can give rise to a higher protest rate. Also, $\frac{\partial^2 p_i^*}{\partial a_i \partial \lambda_i} > 0$ means common aggrievement plays a pivotal role in launching protests given the number of potential protesters. In all, this is how complementarity leads to amplification, which is the key insight of Le Bon (1897).

A.3 Agglomeration Effect or Spillover Effect

Before HSR construction, a representative individual in city i can gain $c_{ii} = w_i - R_i$ or $c_{i0} = w_0 - R_0 - n_i T_i^0$, where T_i^0 is the mobility cost via traditional means of transportation (usually buses or trains). An individual would not migrate so long as the utility difference does not exceed a certain threshold, η_{ij}, which captures the idiosyncratic preference for their hometown. η_{ij} follows a distribution $G_i(\cdot)$, which can be interpreted as a common hometown attachment or the city's stickness. Thus, his probability for migration is given by

$$\rho_i^0 = Pr(u_0 - u_i \geq \eta_{ij}) = G_i(u(w_0 - R_0 - n_i T_i^0) - u(w_i - R_i)).$$

We assume $w_0 - R_0 > w_i - R_i$, so $\rho_i^0 > 0$ at least for some $i \in \mathcal{N}$.[4] In equilibrium, there remains $(1 - \rho_i^0\alpha)N_i$ residents in city i, and $N_0 + \sum_{i=1}^n \rho_i^0 \alpha N_i$ residents in the capital (city 0). For a protest occurring in city i, its potential protesters are as many as $\lambda_i^0 = N_0 + \sum_{i=1}^n \rho_i^0 \alpha N_i + \sum_{|d_k - d_i| \leq r}(1 - \rho_k^0\alpha)N_k$ if $d_i \leq r$, or $\lambda_i^0 = \sum_{|d_k - d_i| \leq r}(1 - \rho_k^0\alpha)N_k$ if $r < d_i \leq D$.

HSR provides the residents with a new approach to reaching the capital. Denote $T_i^1(d_i)$ as

4. This also helps to prevent "adverse migration": residents in the center will not migrate to other prefectures. Such simplification is proper as we focus on economically-driven migration that naturally favor the center.

the mobility cost via HSR. According to conventions in economic geography (See e.g., Pan et al. (2023)), HSR is more economical than buses or trains when the distance is far (above a threshold d^*).[5] Thus, we assume $(T_i^1(d_i) - T_i^0(d_i))(d_i - d^*) \leq 0, \forall d_i \in [0, D]$, and only those living in remote areas ($i \in A_{d^*} = \{i : d^* < d_i \leq D\}$) would surely choose HSR in lieu of buses or trains. For these groups of people, their probability for migration is altered to

$$\rho_i^1 = Pr(u_0 - u_i \geq \eta_{ij}) = G_i(u(w_0 - R_0 - n_i T_i^1) - u(w_i - R_i)).$$

Since both G_i and u are increasing functions, $T_i^1(d_i) < T_i^0(d_i), \forall d_i \in A_{d^*}$, it is evident that $\rho_i^1 > \rho_i^0$. Hence, the number of potential protesters in city i is $\lambda_i^1 =$

$$\begin{cases} N_0 + \sum_{d_i \leq d^*} \rho_i^0 \alpha N_i + \sum_{d_i > d^*} \rho_i^1 \alpha N_i + \sum_{|d_k - d_i| \leq r} (1 - \rho_k^0 \alpha) N_k > \lambda_i^0, & \forall d_i \leq r \\ \sum_{|d_k - d_i| \leq r} (1 - \rho_k^0 \alpha) N_k = \lambda_i^0, & \forall r < d_i \leq d^* \\ \sum_{|d_k - d_i| \leq r} (1 - \rho_k^1 \alpha) N_k < \lambda_i^0, & \forall d^* < d_i \leq D \end{cases}$$

Given Equation A2, $\frac{\partial p_i^*}{\partial \lambda_i} > 0$, we have the following proposition:

Proposition A1 (Agglomeration Effect) *HSR opening can foster migration, thus*

- *increasing protests in central areas by attracting more migrants ($\lambda_i^1 > \lambda_i^0$);*
- *having little (ambiguous) influence for regions ($\lambda_i^1 \approx \lambda_i^0$);*
- *reducing protests in remote areas by pushing residents away ($\lambda_i^1 < \lambda_i^0$).*

The above analysis shows how agglomeration contributes to HSR's influence. We further take into consideration the potential spillover in the long run. In spatial equilibrium, individuals should be indifferent between migration or staying in their hometown. With the decrease in transportation cost ($T_i^1(d_i) \leq T_i^0(d_i)$), $w_i - R_i$ should increase, with a larger rise in more distant regions. Alternatively, $w_0 - R_0$ should decrease. In the former case, people gain a much higher surplus ($c_i^1 > c_i^0$), thus lower aggrievement ($a_i^1 = a(\bar{u} - u(c_i^1)) < a(\bar{u} - u(c_i^0)) = a_i^0$). Hence, they are less willing to participate in protests both in aggregation and in margin ($p_i^1 < p_i^0, \frac{\partial p_i^1}{\partial \lambda_i} < \frac{\partial p_i^0}{\partial \lambda_i}$). The latter is just the reverse, with reduced utility and more protests in central regions.

Proposition A2 (Spillover Effect) *HSR opening can foster resource spillover, thus*

- *increasing protests for central areas with lower wages or higher housing prices;*
- *reducing protests in remote areas with higher wages or lower housing prices.*

5. We assume $d^* > r$ since a protest could hardly mobilize people far from the activity center. In particular, people living in remote areas would hardly participate in protests occurring in central regions, as shown by our empirical evidence later. We roughly estimate $d^* \in [120, 160]$ in China while most literature shows $r < 20$.

Analogous results can be obtained by considering the general equilibrium of $n+1$ labor markets and $n+1$ housing markets. For instance, more migrants provide the central area with more labor supply and housing demand, giving rise to R_0 while reducing w_0. With smaller $w_0 - R_0$, the migrant's utility $u(w_0 - R_0 - n_i T_i)$ is reduced, leading to higher aggrievement and higher protest rate. Nevertheless, such an analysis would bring about excessive words and unnecessary difficulties in illuminating our key results, so we choose a simpler version.

A.4 Monitoring Effect or Deterrence Effect

Aside from its economic implications, HSR acts as an instrument for increasing state presence. The provincial capital is definitely the political center, and the provincial officials manage to project power to their subordinate regions. HSR network reduces such cost, endowing them with greater capacity both in monitoring local corruption and in repressing local protests, enhancing repression.

To account for the influence of local corruption, we denote local corruption rate as q_i^c and assume individuals are more aggrieved when higher corruption. Thus, $a_{ij} = a(\bar{u} - u_{ij}, q_i^c)$ is a function of utility gap and local corruption, increasing in both terms. Alternatively, q_i^c can be regarded as the rent-seeking proportion, like excessive tax or fees, and $a_{ij} = a(\bar{u} - u(c_{ij}(1 - q_i^c)))$, which also satisfies the above conditions. Consider an inspector a choosing the optimal city for anti-corruption.Her preference is shaped by two major factors, local corruption rate (q_i^c), and the cost to reach there ($T_i^x, x = 0, 1$). By conducting inspections in more corrupt cities, her probability of detecting (and eradicating) corruption is higher, which not only fulfills her commission but also acts as her achievements that facilitate her future promotion. Her utility is written by $V_{ia} = (\phi q_i^c - T_i^x(d_i))\sigma_{ia}$, where σ_{ia} captures the inspector's personal responsibility and her preference for particular regions (for instance, her own hometown). Following conventions in modeling city choices, we assume σ_{ia} follows the Frechet distribution $F_a(\sigma_{ia}) = exp(-L_a \sigma_{ia}^{-\beta})$, where $L_a > 0$ is the level and $\beta > 1$ is the dispersion. Therefore, before HSR opening, the probability of being inspected for officials in city i's (the inspection rate) is given by

$$q_i^I = Pr(V_{ia} = \max_{k \in \{1,\dots,n\}} V_{ka}) = \frac{(\phi q_i^c - T_i^0(d_i))^\beta}{\sum_{k \in \{1,\dots,n\}} (\phi q_k^c - T_k^0(d_k))^\beta}.$$

For a local government official (e.g., the mayor or party secretary), her optimal corruption rate q_i^C should be as high as possible without incurring inspection. Once she becomes highly corrupted vis-à-vis her counterparts (officials in other cities), she would be inspected, put off, and even accused. Such disutility is too low for her to bear, which facilitates them to compete to reduce the corruption rate like a Bertrand game. In equilibrium, the inspection rate should be equalized to $q_i^{I0} = \frac{1}{n+1}$ and $q_i^{C0} = T_i^0(d_i)/\phi + q_0^{C0}$. After HSR opening, the trans-

portation cost is reduced for remote areas, which also reduces the equilibrium corruption rate: $q_i^{C1} = T_i^1(d_i)/\phi + q_0^{C1} < T_i^0(d_i)/\phi + q_0^{C0} = q_i^{I0}$ if $d^* < d_i \leq D$. In consequence, residents there develop lower aggrievement ($a_i^1 < a_i^0$), reducing both the equilibrium protest rate and marginal protest occurrence ($p_i^1 < p_i^0$, $\frac{\partial p_i^1}{\partial \lambda_i} < \frac{\partial p_i^0}{\partial \lambda_i}$).

Proposition A3 (Monitoring effect) *HSR opening can facilitate investigations into local officials, alleviating corruption and reducing protests in remote areas.*

Similarly, the public security leader b is also choosing the city for anti-protest. Her preference is shaped by local protest rate p_i and the difficulty in reaching there $T_i^x(x = 0, 1)$, written by $V_{ib} = (\psi p_i - T_i^x(d_i))\sigma_{ib}$, with σ_{ib} capturing her idiosyncratic characteristics. We still assume σ_{ib} following the Frechet distribution $F_b(\sigma_{ib}) = exp(-L_b\sigma_{ia}^{-\gamma})$, where $L_b > 0$ is the level and $\gamma > 1$ is the dispersion. Before HSR opening, the repression probability of a protest in city i is given by

$$q_i^R = Pr(V_{ib} = \max_{k\in\{1,\ldots,n\}} V_{kb}) = \frac{(\psi p_i - T_i^0(d_i))^\gamma}{\sum_{k\in\{1,\ldots,n\}}(\psi p_k - T_k^0(d_k))^\gamma}.$$

As the cost of launching protests hinges directly upon the repression rate, we can substitute it into Equation A1: $p_i = F_i(p_i\lambda_i a_i - \mu - \mu_i(q_i^R))$. Since $\mu_i^{'} > 0$ and $\frac{\partial q_i^R}{\partial T_i} < 0$, with reduced transportation cost by HSR opening ($T_i^1(d_i) < T_i^0(d_i)$ for $d^* < d_i \leq D$), the repression rate (along with the cost of protest) is increased. In other words, residents are deterred by increasing state presence. For fear of being caught to jail, they would be more prudent and the protest rate is reduced.

Proposition A4 (Deterrence Effect) *HSR opening can extend state presence, deterring residents from launching protests, thus reducing protests in remote areas.*

A.5 Proof for existence and uniqueness

To ensure the existence of the equilibrium, we must assume there is a positive mass of individuals who are willing to engage in protests even if they expect to be alone ($F_i(-\mu - \mu_i) > 0$), and a positive mass who never participates even if they expect the whole group to join the resistance ($F_i(\lambda_i a_i - \mu - \mu_i) < 1$). The former corresponds to the "initiators" of the resistance. As mentioned in Granovetter (1978), they are group loyalists who launch the protest and mobilize other members to participate. The latter consists of passive members who would never engage in resistance, like those civil servants who should always support the government.

Assumption A1 $F_i(-\mu - \mu_i) > 0$ *and* $F_i(\lambda_i a_i - \mu - \mu_i) < 1$.

Given the complementarity, multiple equilibria are possible. To rule out multiplicity, we assume

enough heterogeneity within the group, at least in a neighborhood of the equilibrium participation rate p_i^* (following Passarelli and Tabellini (2017)).

Assumption A2 $\lambda_i a_i f_i(p_i \lambda_i a_i - \mu) < 1$.

Lemma A1 *Assumption A1 is necessary and sufficient for the existence of an equilibrium participation rate* $0 < p_i^* < 1$. *Both Assumption A1 and Assumption A2 are necessary and sufficient for the uniqueness of* p_i^*.

Proof. The inequalities in Assumption A1 ensure that the fixed-point(s) are not on the boundaries, i.e., $p_i^* > 0$ and $p_i^* < 1$, respectively. In addition, thanks to continuity of F_i, Equation A1 has at least one solution in $(0, 1)$, which is a direct application of the Brouwer's fixed-point theorem. Assumption A2 requires that F_i crosses the p_i line from the above side under all circumstances. Since F_i is continuous, it follows that the solution must be unique.

For necessity, observe that $\lim_{p_i \to -\infty} F_i = 0$ and $\lim_{p_i \to \infty} F_i = 1$. If there exist $p_i^* \in (0, 1)$, neither $p_i = 0$ nor $p_i = 1$ can be a solution. This implies Assumption A1. Since the solution lies in $(0, 1)$, F_i crosses p_i from above at least once. If p_i^* is unique, then the cross should be once, which implies Assumption A2. See Passarelli and Tabellini (2017) for detailed proofs. ■

Appendix B More on Background and Data

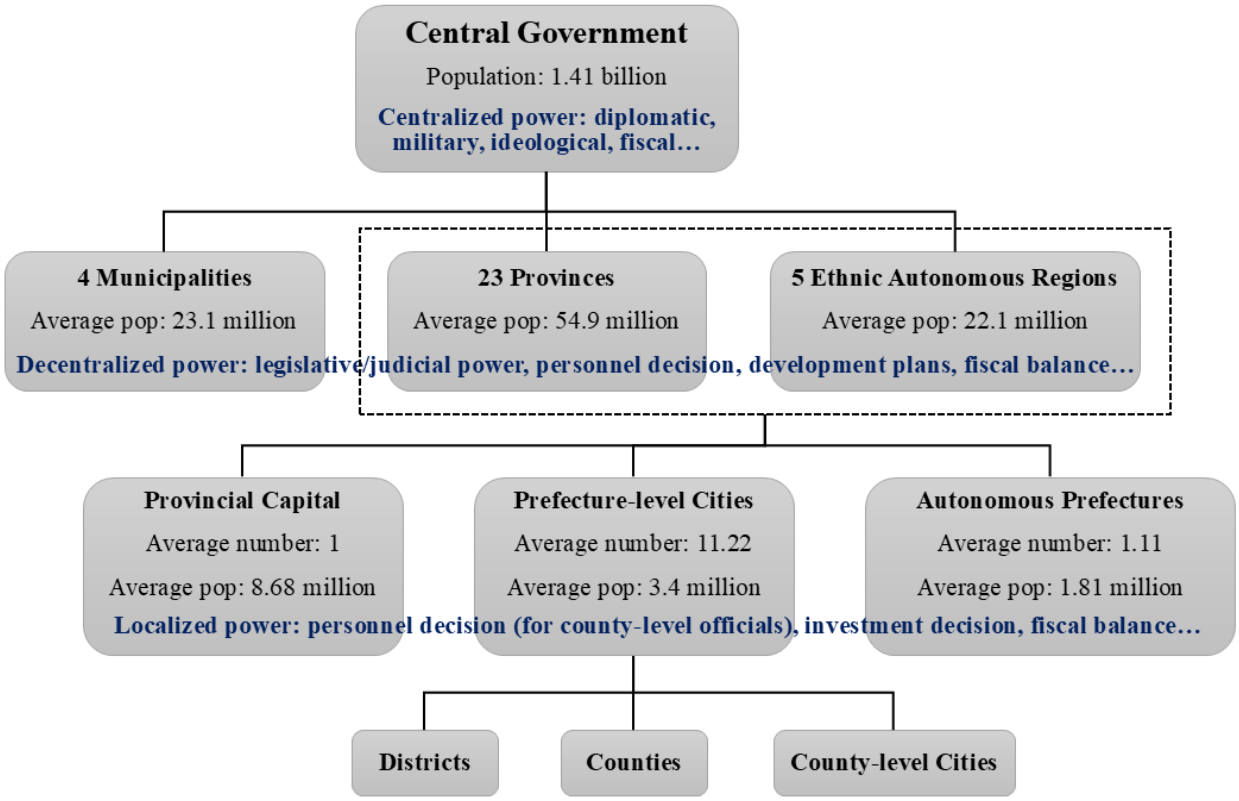


FIGURE B1 The Divisions of Chinese Bureaucracy.

Notes: This figure presents the administrative divisions in China and the relations between different divisions. From top to bottom is: the central level, the province level, the prefecture level, and the county level. The lowest level (town-level) is abbreviated for clarity. Our sample includes 22 provinces and 5 ethnic autonomous regions (consisting of provincial capitals, prefecture-level cities, and autonomous prefectures).

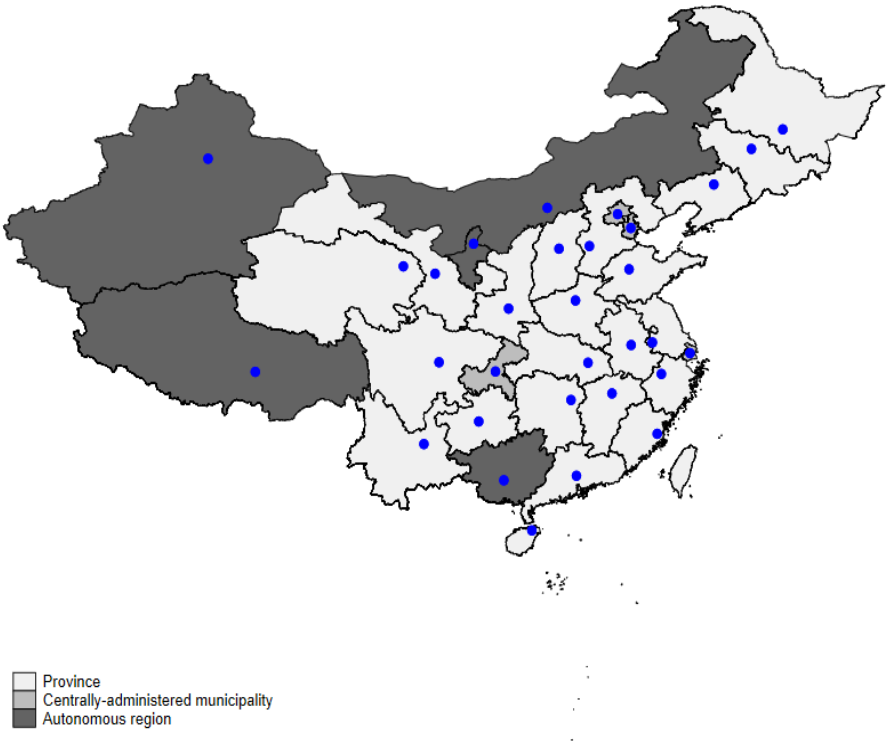


FIGURE B2 The Spatial Distribution of Provincial Regions.

Notes: This figure presents the spatial distribution of Chinese provinces and provincial capitals. The blue dots represent the political center (provincial capital) of each province.

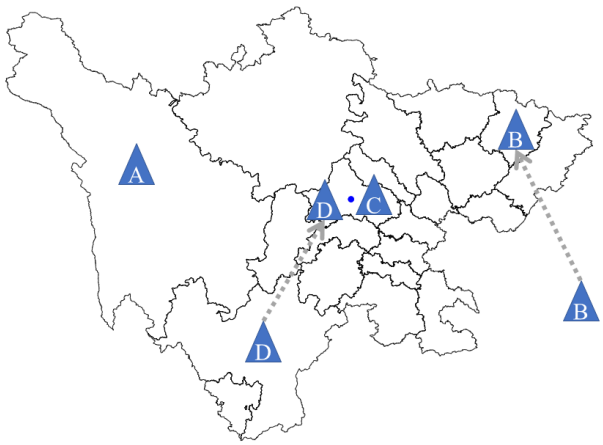


FIGURE B3 A sketch of different groups of people.

Notes: This figure illustrates four types of individuals discussed in the paper, using Sichuan Province as an example. The blue dot marks the provincial capital, Chengdu. Group A refers to local residents in peripheral (remote) areas. Group B consists of migrants who move to peripheral areas. Group C represents local residents in central regions. Group D includes migrants relocating to central regions. Group D (A-to-NA migrants) is our main focus, contributing to increased protests in central regions following HSR opening. Group A is subject to the deterrence effect; Group C may benefit from the agglomeration effect; Group B is not our focus.

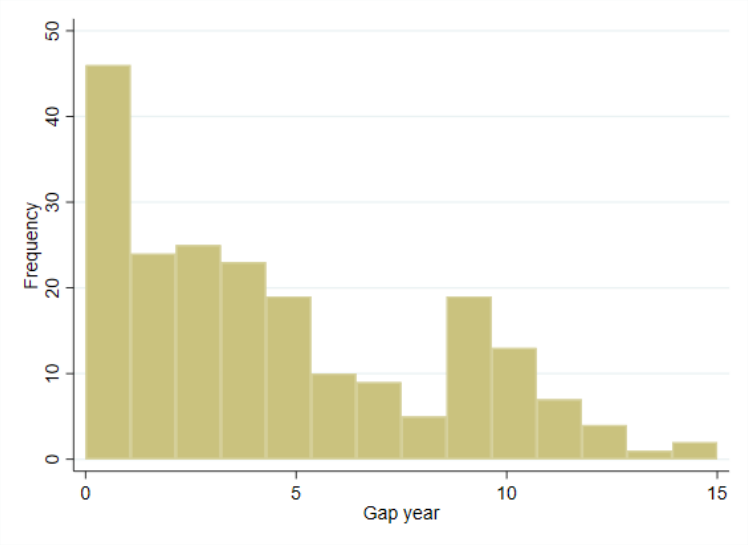


FIGURE B4 The HSR opening time gap between outer regions and capitals.

Notes: This figure presents the different time gap between the opening year of HSR stations and that in their provincial capitals during 2010-2022. Larger year means later HSR opening.

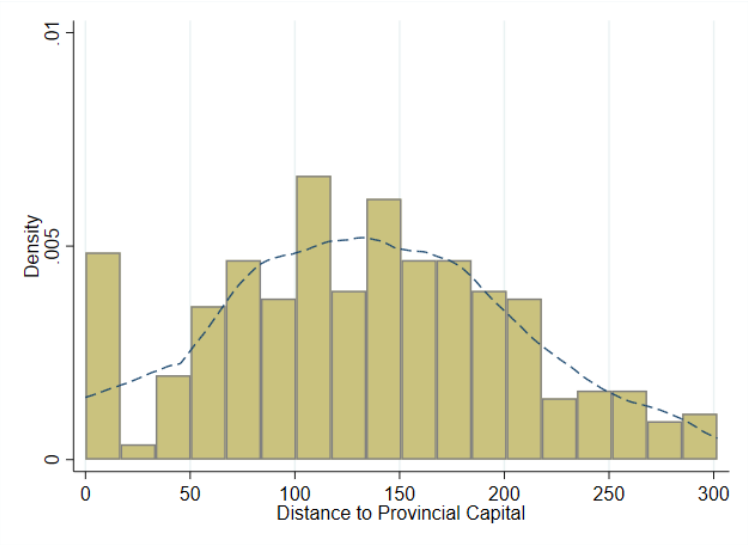
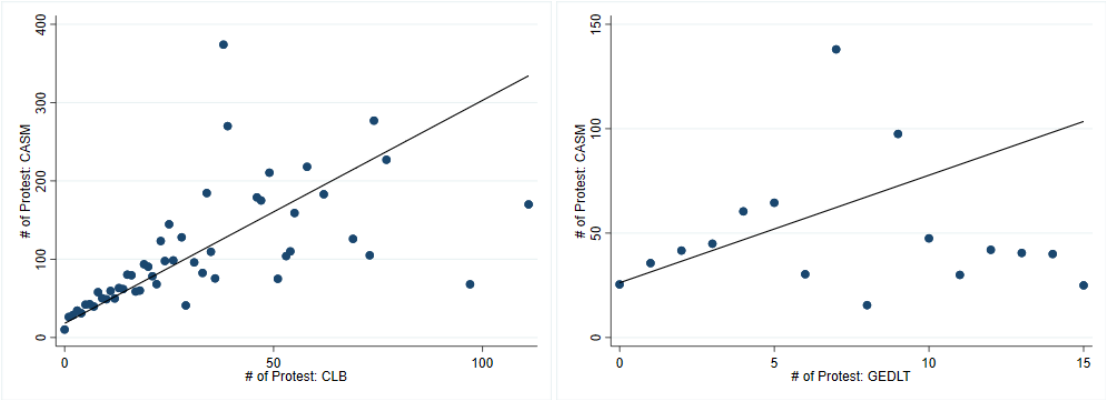


FIGURE B5 The density of geographical distances to provincial capital.

Notes: This figure presents the distribution of the geographical distance between prefectures and the provincial capital in China. The data for municipalities is excluded in our sample.



(A) **Correlation: CASM and CLB.** (B) **Correlation: CASM and GDLET.**

FIGURE B6 The Correlation between the Number of Protests in Different Datasets.

Notes: This figure illustrates the correlation between the number of protests in the CASM, CLB, and GDELT datasets.

TABLE B1 Relationship between HSR and Internet Penetration

	(1)	(2)	(3)	(4)
Panel A: Dependent variable: *Internet Penetration*				
HSR	-0.032	-0.076	-0.346	-0.333
	(0.052)	(0.053)	(0.304)	(0.328)
HSR×*lndistance*			0.074	0.060
			(0.062)	(0.067)
Observations	1,902	1,895	1,889	1,882
R-squared	0.782	0.829	0.783	0.829
Panel B: Dependent variable: *Weibo Total Count*				
HSR	0.053	0.058	0.129	0.058
	(0.075)	(0.161)	(0.178)	(0.161)
HSR×*lndistance*			-0.022	0.002
			(0.033)	(0.030)
Observations	1,080	1,076	1,072	1,068
R-squared	0.763	0.820	0.762	0.819
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Year FE	✓		✓	
Province-year FE		✓		✓

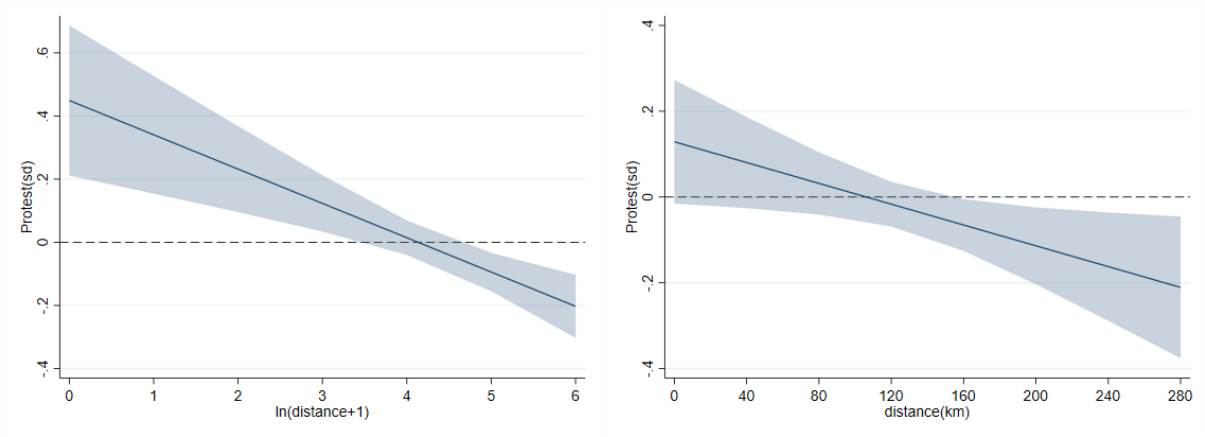
Notes: Table B1 reports the relationship between HSR access and internet penetration. The dependent variable is the standardized number of households with internet access in each prefecture (*Internet Penetration*), with data covering the years 2010–2016. Dependent variable in Panel B is the standardized total number of *Weibo* posts (*Weibo Total Count*). The data for Panel B spans the years 2010–2013. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. Robust standard errors in parentheses are clustered at the prefecture levels. ***p< 0.01, **p <0.05, *p < 0.1.

TABLE B2 Summary Statistics: CLB Dataset

	(1)	(2)	(3)	(4)	(5)
Variable	Obs	Mean	Std.Dev.	Min	Max
Panel A: Independent Variables					
HSR	2,584	0.464	0.499	0	1
Panel B: Prefecture-year Outcomes[a]					
Strike	2,584	3.757	7.514	0	111
Strike(sd)	2,584	0	1	-0.500	14.273

Notes: Table B2 presents descriptive statistics for our main variables using CLB dataset.

Appendix C More on Empirical Results

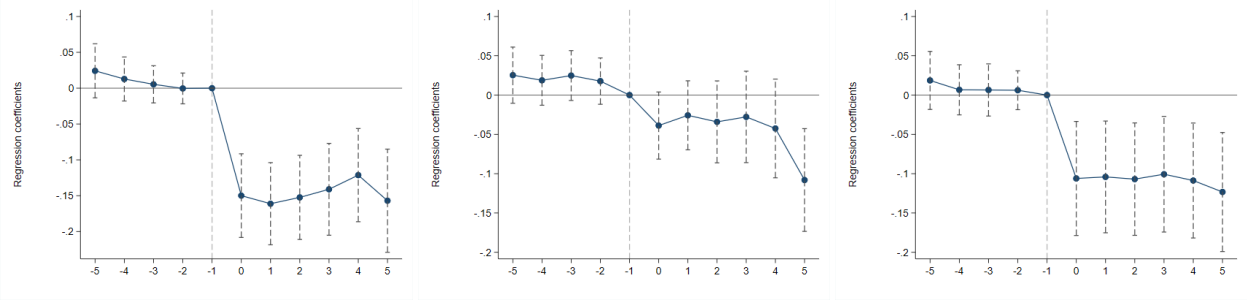


(A) **The heterogeneous impact for different distance (logged) from provincial capital.**

(B) **The heterogeneous impact for different actual distance from provincial capital.**

FIGURE C1 The heterogeneous impact of HSR opening with variation in distances to the provincial capital.

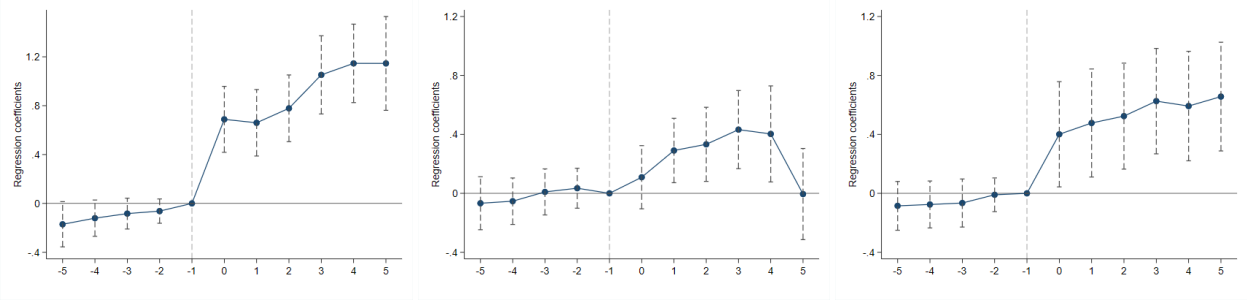
Notes: This figure plots the heterogeneous impact of HSR opening on residents with varied distances to the provincial capital. In Panel A, we use the distance (logged) to capture heterogeneity. In Panel B, we replace it with actual distance (per kilometer) and re-estimate the result. The blue line represents the impact of HSR opening on the number of protests at different distances. The dependent variable is the standardized number of protests at the prefecture level (*Protestsd*). Both panels control for Province-year fixed effects, predetermined-year and prefecture fixed effects. The shaded blue area represents the 95% confidence interval.



(A) **Convention events.** (B) **Disruptive events.** (C) **Violent events.**

FIGURE C2 Pre-trend tests for different protest intensity (Interaction term)

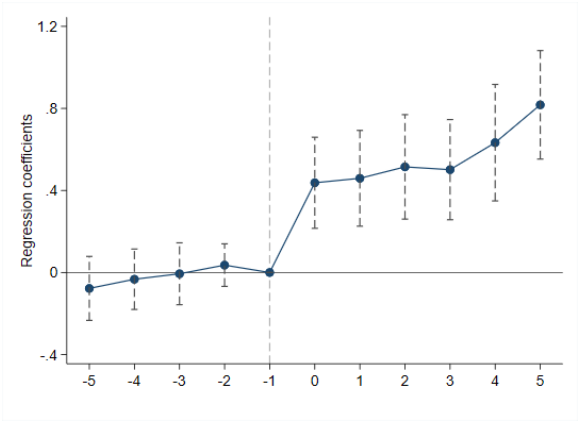
Notes: This figure plots pre-trend tests for interaction term *HSR*×*lndistance* in equation (2) using different classifications based on intensity. The dependent variables for Panels A-C are standardized prefecture-level protest counts of different intensities, namely conventional, disruptive, and violent protests. All panels control for Province-year fixed effects, predetermined-year and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.



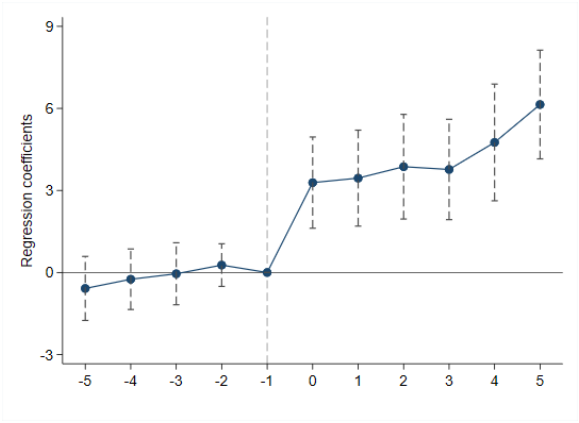
(A) **Convention events.** (B) **Disruptive events.** (C) **Violent events.**

FIGURE C3 Pre-trend tests for different protest intensity (HSR term)

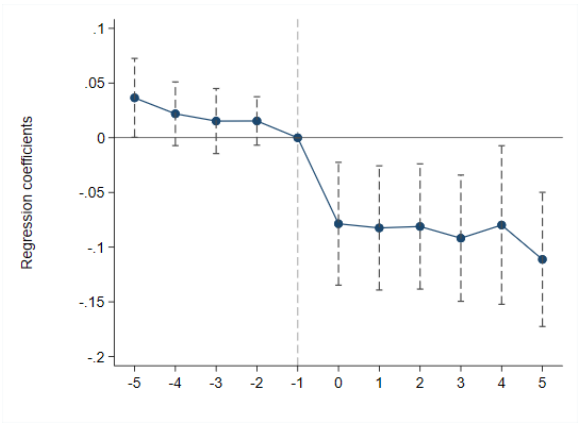
Notes: This figure plots pre-trend tests for single term *HSR* in equation (2) using different classifications based on intensity. The dependent variables for Panels A-C are standardized prefecture-level protest counts of different intensities, namely conventional, disruptive, and violent protests. All panels control for Province-year fixed effects, predetermined-year and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.



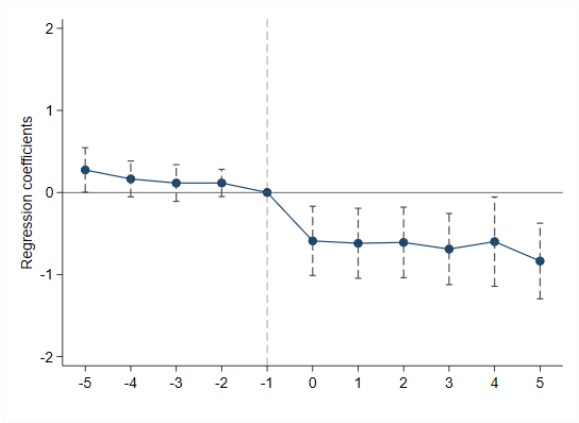
(A) **Pre-trend tests for *HSR* term using the standardized number of protests.**



(B) **Pre-trend tests for *HSR* term using the number of protests.**



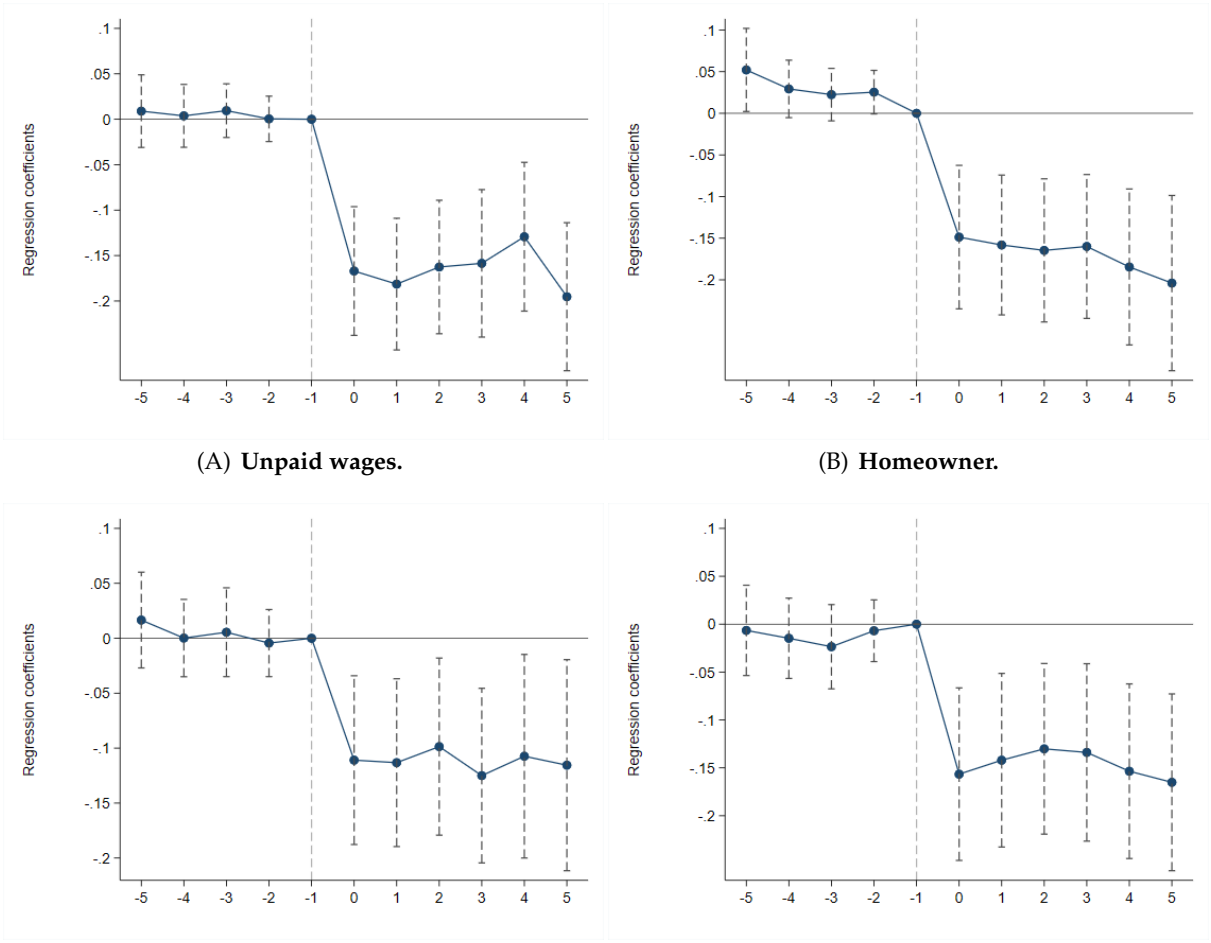
(C) **Pre-trend tests for interaction term using the standardized number of protests.**



(D) **Pre-trend tests for interaction term using the number of protests.**

FIGURE C4 Pre-trend tests using CLB dataset.

Notes: This figure plots pre-trend tests using prefecture-year level CLB dataset. The dependent variable is the standardized number of strikes at the prefecture level in Panel A and Panel C, while we use the actual number of protests (*Protest*) in Panel B and Panel D. Panel A and Panel B test for single *HSR* term. Panel C and Panel D test for the interaction term *HSR*×*lndistance* in equation (2); All panels control for province-year, predetermined-year, and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.

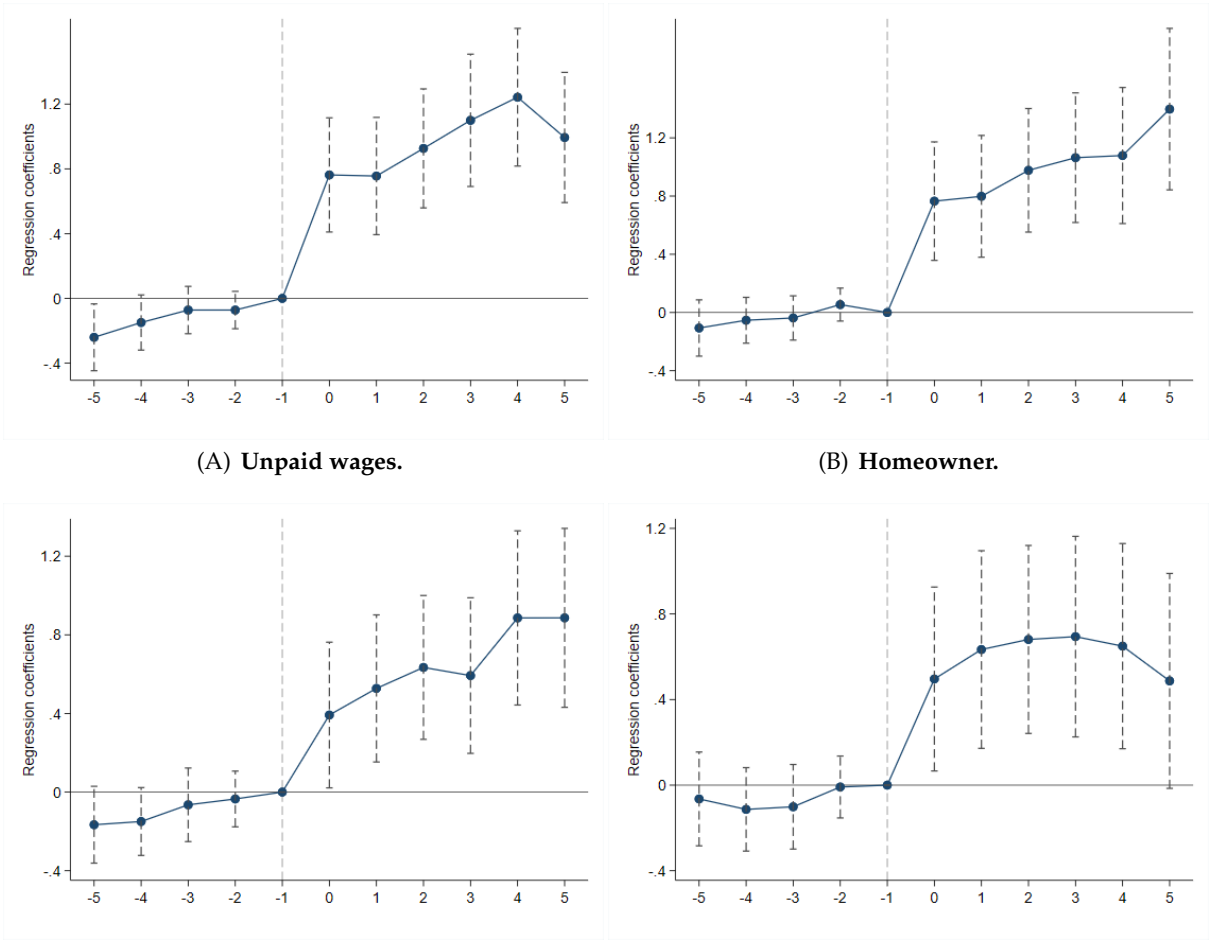


(A) **Unpaid wages.** (B) **Homeowner.**

(C) **Education.** (D) **Medicine.**

FIGURE C5 Pre-trend tests for different protest issues (Interaction term).

Notes: This figure plots pre-trend tests for interaction term *HSR*×*lndistance* in equation (2) using different kinds of protests based on motives. The dependent variables for Panels A-D are standardized prefecture-level protest counts associated with unpaid wages, homeowner, education, and medicine, respectively. All panels control for Province-year fixed effects, predetermined-year and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.



(A) **Unpaid wages.** (B) **Homeowner.**

(C) **Education.** (D) **Medicine.**

FIGURE C6 Pre-trend tests for different protest issues (*HSR* term).

Notes: This figure plots pre-trend tests for *HSR* in equation (2) using different kinds of protests based on contents. The dependent variables for Panels A-D are standardized prefecture-level protest counts associated with unpaid wages, homeowner, education, and medicine, respectively. All panels control for Province-year fixed effects, predetermined-year and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.

TABLE C1 Baseline using different distance bins.

	(1)	(2)
HSR	0.523***	0.524***
	(0.141)	(0.141)
	[0.142]	[0.142]
HSR×*Segment2*	-0.552***	-0.568***
	(0.142)	(0.148)
	[0.144]	[0.147]
HSR×*Segment3*	-0.654***	-0.601***
	(0.141)	(0.139)
	[0.138]	[0.139]
HSR×*Segment4*	-0.567***	-0.550***
	(0.150)	(0.151)
	[0.145]	[0.141]
Predetermined-year Control	✓	✓
Prefecture FE	✓	✓
Province-year FE	✓	✓
Observations	1,896	1,896
R-squared	0.900	0.900

Notes: Table C1 reports baseline regression results for different distance bins. The dependent variables are the prefecture level standardized counts of protests. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. Column (1) examines the effect of HSR on protest using distance quartiles between prefectures and their provincial capitals. Column (2) uses alternative distance bins (0–20 km, 20–100 km, 100–200 km, and over 200 km) to assess HSR's heterogeneous effect. Robust standard errors in parentheses are clustered at the prefecture levels. Standard errors in brackets are Conley standard errors robust for spatial correlation, assuming a cutoff window of 500 km and a serial correlation of 6 years. ***p< 0.01, **p <0.05, *p < 0.1.

TABLE C2 Baseline Results using Heterogeneous Robust DID estimator

(1) Group	(2) Estimate	(3) SE	(4) LB CI	(5) UB CI
Full sample	0.092	0.059	-0.024	0.208
Closest	0.620	0.225	0.179	1.062
Near	-0.049	0.050	-0.146	0.048
Mid	-0.102	0.083	-0.265	0.061
Far	-0.129	0.091	-0.307	0.050

Notes: Table C2 reports baseline regression results using Heterogeneous Robust DID estimator from De Chaisemartin and d'Haultfoeuille (2020). Robust standard errors are clustered at the prefecture levels. The dependent variable is the standardized number of protests at the prefecture-year level (*Protestsd*). The first row reports the impact coefficient of high-speed rail opening (*HSR*) on the standardized protest frequency for the full sample. The second to fourth rows report the sub-sample results, from closest to farthest from the provincial capital, for the impact coefficient of *HSR* on *Protestsd*. The stratification is based on the prefecture's distance from the provincial capital, divided equally into four quartiles. To follow the heterogeneity robust estimates of De Chaisemartin and d'Haultfoeuille (2020), we excluded prefectures that are always treated. Column (3) reports the standard errors of the estimations, while columns (4)-(5) report the lower and upper bounds of the 95% confidence interval, respectively.

TABLE C3 Classification Based on Protest Intensity

	(1) Baseline	(2) Convention	(3) Disruptive	(4) Violent
HSR	0.467***	0.649***	0.180	0.436**
	(0.134)	(0.146)	(0.113)	(0.185)
	[0.126]	[0.145]	[0.123]	[0.436]
HSR×*lndistance*	-0.109***	-0.146***	-0.042*	-0.106***
	(0.027)	(0.030)	(0.023)	(0.037)
	[0.025]	[0.031]	[0.024]	[0.035]
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,896	1,896	1,896	1,896
R-squared	0.900	0.855	0.821	0.838

Notes: Table C3 reports baseline regression results for different classifications based on protest intensity. Column (1) reports the baseline results. The dependent variables in columns (2) to (4) are the prefecture level standardized counts of protests of different intensities, including conventional, disruptive, and violent protests. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. Robust standard errors in parentheses are clustered at the prefecture levels. Standard errors in brackets are Conley standard errors robust for spatial correlation, assuming a cut-off window of 500 km and a serial correlation of 6 years. ***p< 0.01, **p <0.05, *p < 0.1.

Appendix D More on Robustness Checks

D.1 Placebo Test

In this section, we strengthen our conclusion through several placebo tests.

First, to assess the plausibility of using distance to the provincial capital as a key variable, we test the impact of other distances in Table D1. In column (1), we replace the distance to the provincial capital by the distance to Beijing, capital of China. If Beijing could exert direct control over prefectures, or if there were a predominant flow of people toward Beijing, then the coefficient of the interaction term between the distance to Beijing and HSR opening should be significant. The result indicates that this effect is probably non-existent. In columns (2)-(4), we test the influence of other national economic centers (Shanghai and Guangzhou), with similar insignificant results. In column (5), we replace the original distance with the distance to the nearest provincial capital, and in column (6), with the distance to the largest non-capital city within a province. Both specifications capture possible economic attraction without political power, and we find little effect. Combined together, the spatial pattern of protest emerges only when using the distance to provincial capital, which supports our focus on within-province outcomes. Specifically, the capital's economic centripetal power and political repression force jointly shape the protest geography.

Second, we use historical variables for placebo tests in Table D2. In column (1), we employ the administrative divisions of Qing Dynasty to measure the distance from a prefecture to the Qing provincial capital. The coefficient of column (1) reflects a similar pattern to our baseline but is not as significant. This demonstrates that changes in the spatial distribution of protests caused by HSR are attributed mainly to the modern economic and political geography rather than historical ones. Additionally, in columns (2)-(4), we use distances to major cities in the Ming and Qing dynasties, silk trade cities, and tea trade centers as placebos, with data derived from Chen, Kung, and Ma (2020). The coefficients are all insignificant, corroborating our findings in the baseline.

Finally, we randomly permute the opening year of HSR across different prefectures to check whether our results are affected by omitted variables. We calculate the "pseudo effect" based on Equation 2 and repeat this process 200 times to improve the accuracy of the placebo test. The results in Figure D1 indicate that our baseline results are not significantly affected by this issue.

TABLE D1 Placebo test: distances to other economic centers

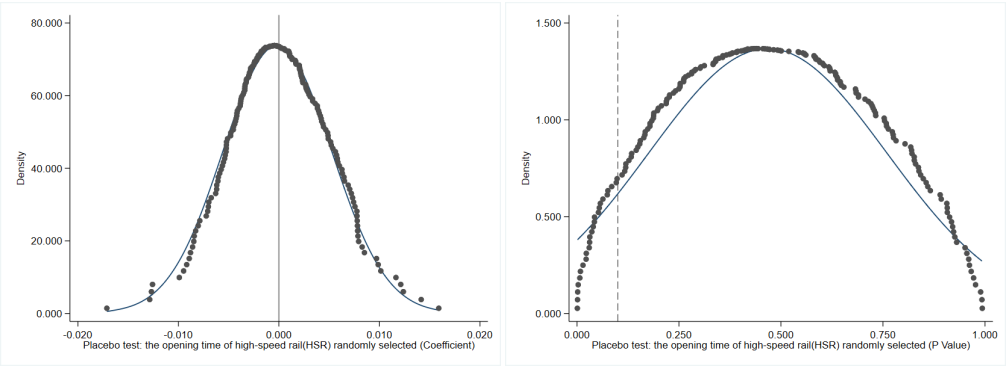
	(1)	(2)	(3)	(4)	(5)	(6)
HSR	-0.382	-0.275	0.440	0.609	0.068	0.051
	(0.588)	(0.546)	(0.429)	(0.433)	(0.078)	(0.171)
HSR×lndistance(Beijing)	0.052					
	(0.083)					
HSR×lndistance(Shanghai)		0.037				
		(0.078)				
HSR×lndistance(Guangzhou)			-0.067			
			(0.061)			
HSR×lndistance(min BSG)				-0.099		
				(0.066)		
HSR×lndistance(min capital)					-0.001*	
					(0.000)	
HSR×lndistance(min largest city)						-0.015
						(0.032)
Predetermined-year Control	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓
Observations	1,910	1,910	1,910	1,910	1,910	1,910
R-squared	0.897	0.897	0.897	0.897	0.897	0.897

Notes: Table D1 reports results for a placebo test using distance measures to other economic centers. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance(Beijing)*, *lndistance(Shanghai)*, *lndistance(Guangzhou)* represent the distance (logged) between the prefecture center and Beijing, Shanghai, Guangzhou, respectively. *lndistance(min BSG)* represents the logarithm of the shortest distance between the prefecture center and three megacities. *lndistance(min capital)* represents the logarithm of the shortest distance between the prefecture center and all provincial capitals in China, while *lndistance(min largest city)* represents the logarithm of the shortest distance between the prefecture center and the largest non-capital city within a province.

TABLE D2 Placebo Test: distances to historical centers

	(1)	(2)	(3)	(4)
HSR	0.811	-0.482	-0.096	-0.138
	(0.921)	(0.587)	(0.085)	(0.110)
HSR×lndistance(Qing)	-0.069			
	(0.076)			
HSR×lndistance(Large Cities)		0.040		
		(0.051)		
HSR×lndistance(Silk)			0.007	
			(0.007)	
HSR×lndistance(Tea)				0.010
				(0.008)
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,616	1,616	1,616	1,616
R-squared	0.901	0.901	0.901	0.901

Notes: Table D2 reports results for a placebo test using distance measures to historical centers. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance(Qing)*, *lndistance(Large Cities)*, *lndistance(Silk)*, *lndistance(Tea)* represent the logarithmic distance of the prefecture to the provincial capital during the Qing Dynasty, the nearest major metropolitan area in 1920, the nearest silk trade center, and the nearest tea trade center, respectively.



(A) **Placebo test for coefficients.** (B) **Placebo test for p-values.**

FIGURE D1 Placebo test: randomly setting HSR opening.

Notes: This figure reports the coefficient distribution of "pseudo change effect" of the opening time of HSR (HSR) estimated by equation (2) after randomly setting the opening year. The dependent variable is the standardized number of protests (*Protestsd*). This process was repeated 200 times so as to improve the confidence of the placebo test. Panel A reports the distribution of coefficients from the randomization permutation test for the interaction terms (*HSR×lndistance*). Panel B reports the corresponding p-values. Both panels control for province-year, predetermined-year, and prefecture-level fixed effects. Robust standard errors are clustered at the prefecture levels.

D.2 Recenter IV Approach, Long-difference Setting and Market Access Exposure

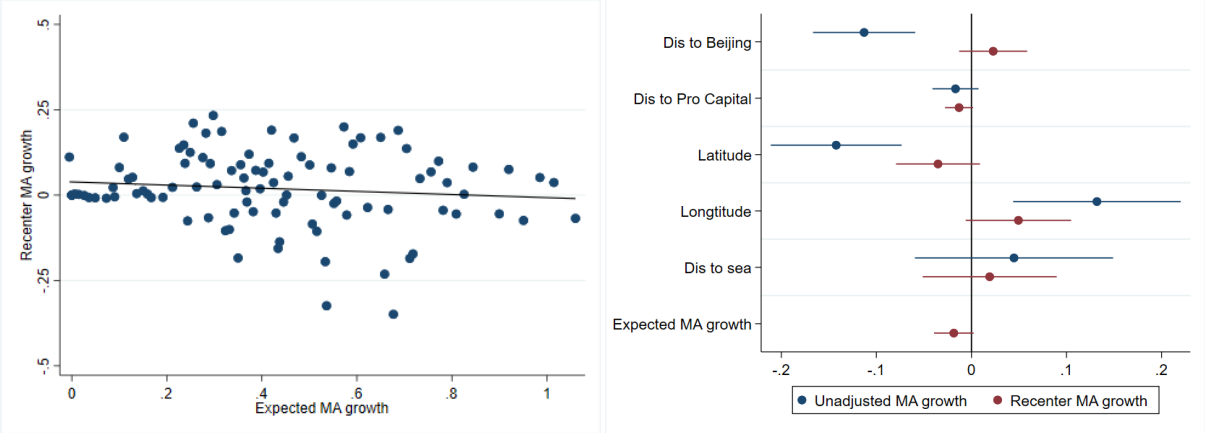
In Table D3, we construct a cross-sectional dataset to examine the relationship between changes in HSR access and changes in protest events during 2010-2017. We restrict the sample to regions without HSR access in 2010 and obtain 226 prefectures (53 of 279 prefectures had HSR in 2010). We then define a dummy variable, ΔHSR, which is 1 if a prefecture has HSR access in 2016 but not in 2010, and 0 otherwise. In columns (1) and (2), we examine how changes in HSR (captured by ΔHSR) affects protest dynamics in different regions. The results are consistent with our baseline findings: changes in HSR opening (from 0 to 1) have brought about significantly more protests in the provincial capital, while the effect wanes as the region moves farther from the capital.

Furthermore, we introduce a continuous measure of HSR exposure based on market access (MA) growth following Borusyak and Hull (2023) and Fang, Li, Wang, and Yang (2025). We calculate MA as: $MA_{i,t} = \sum_{j \neq i} \exp(-0.02\tau_{ij,t}) \cdot pop_{j,2010}$, where $pop_{j,2010}$ is the population of city j in 2010, and $\tau_{ij,t}$ is the predicted travel time from region i to j in year t. For each city pair, the travel time is derived from the minimum-time means among the three types of transportation: high-speed rail (HSR), conventional rail, and highway. We then compute the MA growth in city i from 2010 to 2016 as $(\log MA_{i,2016} - \log MA_{i,2010})$. Compared to a binary indicator, the MA growth measure accounts for the number of connected cities, the economic/political significance of these cities (proxied by population size), and the quality of HSR line (proxied by travel speed). In columns (3) and (4), we use MA growth as an alternative measure of exposure to HSR and obtain similar findings. Column (4) shows that an average MA growth of 1 log point over the period increases the number of protests in provincial capitals by 2.46 s.d., with this effect diminishing by 0.51 s.d. for each doubling of the distance from the provincial capital.

To address possible endogeneity concern, we employ the "recentered-IV" approach and use an expected MA growth derived from a counterfactual HSR network. Specifically, we permute the completion status of built and unbuilt HSR lines in 2016 to generate a counterfactual network, repeat this permutation 1,000 times, and compute the expected MA growth as the average MA growth across these permutations. By using the difference between actual and expected MA growth as as an instrument for MA growth, we can isolate variation unrelated to unobserved confounders and identify its causal effect on protests. Columns (5) and (6) show that an average MA growth of 1 log point over the period increase the number of protests in provincial capitals by 2.54 s.d., with this effect diminishing by 0.52 s.d. for each doubling of the distance from the capital.

To validate the rationality of our expected market access (MA) growth specification, we demonstrate that recentered MA growth is uncorrelated with expected MA growth, as shown in Panel A of Figure D2. Additionally, we quantify the systematic nature of spatial variation in MA growth. In the first regression specification depicted in Figure D2 (indicated by blue dots), the blue regression coefficients illustrate the relationship between unadjusted MA growth and economic geogra-

phy variables, including distance to Beijing, distance to the provincial capital, longitude, latitude, and distance to the coastline. These variables are significant predictors of MA growth, with the exception of distance to the coastline, and the model achieves an R^2 of 0.311. However, when recentered MA growth is used as the dependent variable and expected MA growth is controlled for (indicated by red dots), these geography variables become statistically insignificant, and the R^2 drops to 0.077. This indicates that the recentering process effectively removes the systematic geographic variation in market access.



(A) **The relation between Recenter MA growth and Expected MA growth.**

(B) **Regressions of Market Access Growth on Measures of Economic Geography.**

FIGURE D2 Balance Test for Recenter IV Approach.

Notes: These figures present the balance test for the Recenter IV approach. Panel A illustrates the relationship between Recenter MA growth and Expected MA growth. Panel B reports the coefficients from regressions of Market Access Growth on measures of economic geography. The blue regression coefficients depict the relationship between the dependent variable, unadjusted MA growth, and economic geography variables, including distance to Beijing, distance to the provincial capital, longitude, latitude, and distance to the coastline. In contrast, the red regression coefficients represent the relationship between the dependent variable, recentered MA growth, and economic geography variables, including distance to Beijing, distance to the provincial capital, longitude, latitude, distance to the coastline, and expected MA growth. For ease of interpretation, all independent variables in the regressions have been standardized. The line shows a 95% confidence interval.

TABLE D3 Long-difference setting, market access and recenter correction.

	Dep. variable: std($\#\ of\ Protest_{2016} - \#\ of\ Protest_{2010}$)					
	Baseline		Unadjusted OLS	Controlled OLS	Recentered IV	
	(1)	(2)	(3)	(4)	(5)	(6)
ΔHSR	1.386***	1.153***				
	(0.363)	(0.297)				
$\Delta HSR \times lndistance$	-0.250***	-0.235***				
	(0.076)	(0.061)				
MA Growth			3.837***	2.463	4.490***	2.504**
			(0.571)	(1.493)	(1.063)	(1.076)
MA Growth $\times$*lndistance*			-0.729***	-0.510*	-0.914***	-0.519**
			(0.114)	(0.301)	(0.218)	(0.216)
Expected MA growth				0.173		
				(1.454)		
Expected MA growth $\times$ *lndistance*				-0.031		
				(0.289)		
Control		✓		✓		✓
Observations	226	226	279	279	279	279

Notes: Table D3 reports results using a long-difference setting and testing for recentering correction following Borusyak and Hull (2023). ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Robust standard errors are reported in parentheses. Control variables use values from 2010. The dependent variable is the difference in protest counts between 2016 and 2010, normalized to ensure comparability. Columns (1) and (2) present the baseline results using the long-difference setting. ΔHSR is a dummy variable that equals 1 if a prefecture has HSR access in 2016 but not in 2010, and 0 otherwise. The variable *lndistance* represents the logged distance between the prefecture center and its provincial capital. These columns exclude cities with HSR access in 2010, which allows us to focus on the difference in protests between cities that gained HSR access in 2010-2016 and those that remained unconnected. Columns (3) uses unadjusted market access (MA) growth as an alternative measure of HSR exposure. Column (4) estimates an OLS regression with recentered MA growth as treatment and controlling for expected MA growth given by the same HSR counterfactuals. Columns (5) and (6) employ the recentered instrumental variable (IV) method by permuting the opening status of built and unbuilt HSR lines with the same number of cross-prefecture links, where we instrumented both $MA\ Growth$ and $MA\ Growth \times lndistance$ with the two recentered versions.

D.3 More Controls

In this section, we examine whether our baseline results are threatened by several potential confounding factors.

Firstly, we exclude the possible effect of other official features, including birthplace, age, gender, ethnicity, and education level, which are believed to affect the likelihood of official promotion (Li and Zhou, 2005). We control them respectively in Table D4. The results are similar to our baseline, showing that they can hardly drive our main results. We also control the influence of officials' bargaining power by controlling for those party secretaries being in the upper-level CPC Standing Committee (mentioned by Ma (2022)), with similar results.

Secondly, we exclude the influence of alternative transportation modes, such as highways and air travel, which may also have emerged in the prefectures during the same time period (thus affecting the protest geography in the same pattern). We test their influence in Table D5. Specifically, from column (2) to column (5), we control for the opening of airport, annual airline passenger traffic, annual highway passenger traffic, and all above, respectively. The results suggest that the effect of HSR opening persists despite controlling for other transportation modes.

Thirdly, we address concerns that our measurement of protests may be associated with the development of regional Internet access, which expands simultaneously with HSR. To alleviate the bias caused by the lower Internet penetration in some areas, we control for the Internet penetration rate and China's "Broadband China" strategy to promote Internet access in Table D6. The results indicate that our baseline regression results are not influenced by such factors. We also test the influence of HSR opening on Internet infrastructure later in Table E4, and find weak relation.

TABLE D4 Controlling for official characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline	Controlling for characteristics of the municipal party secretary				Controlling for characteristics of the mayor		
		Provincial CPC Standing Committee	Birthplace	Age	Other	Birthplace	Age	Other
HSR	0.467***	0.468***	0.442***	0.475***	0.509***	0.418***	0.323**	0.308*
	(0.134)	(0.136)	(0.123)	(0.131)	(0.137)	(0.139)	(0.162)	(0.183)
HSR×*lndistance*	-0.109***	-0.109***	-0.095***	-0.103***	-0.111***	-0.102***	-0.078**	-0.073*
	(0.027)	(0.028)	(0.029)	(0.030)	(0.031)	(0.029)	(0.034)	(0.038)
Predetermined-year Control	✓	✓	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1,896	1,628	1,552	1,472	1,733	1,584	1,512	
R-squared	0.900	0.927	0.932	0.933	0.923	0.934	0.937	

Notes: Table D4 reports results controlling for official characteristics. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Column (1) reports the baseline results. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. Columns (2) to (5) gradually control for the individual characteristics of the municipal party secretary (*Shigwei Shuji*) . Column (2) controls whether municipal party secretary also serves as a member of the provincial CPC standing committee. Column (3) introduces fixed effects for the birthplace of the municipal party secretary, while column (4) further controls for age and its square term, and column (5) additionally controls for municipal party secretary's gender, ethnicity, and education level. columns (6) to (8) gradually control for the individual characteristics of the mayor (*Shizhang*) . Column (6) introduces fixed effects for the birthplace of the mayor, while column (7) further controls for age and its square term, and column (8) additionally controls for mayor's gender, ethnicity, and education level.

TABLE D5 Controlling for alternative transportation modes.

	(1)	(2)	(3)	(4)	(5)
		Controlling for			
	Baseline	Airport station	Passenger Traffic (air)	Passenger Traffic (bus)	All
HSR	0.467***	0.443***	0.347**	0.475***	0.352**
	(0.134)	(0.133)	(0.143)	(0.133)	(0.144)
HSR×*lndistance*	-0.109***	-0.104***	-0.088***	-0.110***	-0.085***
	(0.027)	(0.027)	(0.029)	(0.027)	(0.029)
Predetermined-year Control	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓
Observations	1,896	1,896	862	1,892	861
R-squared	0.900	0.901	0.931	0.900	0.932

Notes: Table D5 reports results controlling for alternative transportation modes. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Column (1) reports the baseline results. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. In column (2), we incorporate a dummy variable indicating airport opening, along with its interaction with the distance from the prefecture to the provincial capital. Columns (3) and (4) respectively introduce local civil aviation passenger traffic or highway passenger traffic for each prefecture annually, along with their interactions with distance. Column (5) controls for all the variables mentioned above.

TABLE D6 Controlling for Internet and Social Media Penetration.

	(1)	(2)	(3)	(4)	(5)
		Controlling for			
	Baseline	Internet penetration rate	Broadband policy	Total Weibo post	All
HSR	0.467***	0.439***	0.460***	0.460***	0.418***
	(0.134)	(0.143)	(0.133)	(0.134)	(0.146)
HSR×*lndistance*	-0.109***	-0.103***	-0.108***	-0.108***	-0.100***
	(0.027)	(0.029)	(0.027)	(0.027)	(0.030)
Predetermined-year Control	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓
Observations	1,896	1,882	1,896	1,868	1,854
R-squared	0.900	0.902	0.900	0.903	0.905

Notes: Table D6 reports results controlling for Internet. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Column (1) reports the baseline results. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. Columns (2) to (4) respectively control for the Internet penetration rate, Chinese Broadband policy and total number of Weibo post. Column (5) controls for both of the variables mentioned above.

D.4 Sub-sample Tests: Possible Outliers

A key assumption in our analysis is the overlapping of economic and political center in a single province. Admittedly most of them follow this rule, there remain some exceptions: some have economic centers other than the capital; some have a strong competitor with the capital; some are close to other economic centers that may possess stronger appeal... We discuss these exceptions in detail in Table D7 to verify that the assumption does not restrict the validity of our conclusion.

A few provinces (7/29) have an economic center which is not the provincial capital, denoted by "decentralized". Intuitively, residents in such provinces may prefer to migrate to the economic center. This implies a distinct pattern of HSR's effect on protests, which should be carefully dealt with. We exclude samples from decentralized provinces in column (2), and exclude samples from the economic centers of these provinces in column (3). The results show little difference from the baseline in column (1).

There are five "separately planned cities" in China, including Qingdao, Dalian, Xiamen, Ningbo, and Shenzhen. They possess higher administrative levels (vice-provincial level), enjoy more economic resources, and thus may act as a direct competitor to the provincial capital in attracting migrants. We exclude them in column (4), obtaining similar results compared to the baseline. Meanwhile, many provinces have a sub-center city.[6] Albeit with lower centripetal force compared with the center, they may act as "small regional centers" that attract residents from adjacent prefectures. We exclude the 21 sub-centers in column (5) and get similar findings.

While inter-province migration entail huge costs, many people still migrate to the national (or transregional) economic centers due to their strong centripetal force. Notably, Beijing and Shanghai, the political and economic centers of China, respectively, have the strongest appeal to Chinese people. The four municipalities (provincial-level cities) also perform similar role. Major transregional centers[7], though not as attractive in the national scale, dominate other cities in the large region (spanning different provinces). For instance, Wuhan is the transregional center in central China (including three provinces), dominating other cities within the large region. We exclude samples from regions adjacent to Beijing/Shanghai, municipalities, and major transregional centers in column (6)-(8), respectively. The results are fairly robust.

As shown in Figure B1, there are five ethnic autonomous regions in China. They are generically similar to provinces in many fields, but differ primarily in ruling goals, where social stability is emphasized over economic development. We exclude samples from the ethnic autonomous regions in column (9) and obtain robust results. In general, the result holds robust.

6. There are 21 sub-centers in total, stipulated by the provincial governments. See https://www.ndrc.gov.cn/wsdwhfz/202311/t20231115_1361958.html for detail.

7. These include Shenyang (Northeast), Nanjing (East), Wuhan (Central), Shenzhen (South), Chengdu (Southwest) and Xi'an (Northwest), stipulated by the central government. See http://www.ciudsrc.com/webdiceng.php?id=156634 for detail.

TABLE D7 Sub-sample tests.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
HSR	0.467***	0.460***	0.456***	0.457***	0.384**	0.468***	0.431***	0.463***	0.412**
	(0.134)	(0.129)	(0.138)	(0.133)	(0.158)	(0.137)	(0.152)	(0.147)	(0.171)
HSR×*lndistance*	-0.109***	-0.108***	-0.109***	-0.111***	-0.095***	-0.111***	-0.104***	-0.112***	-0.097***
	(0.027)	(0.026)	(0.028)	(0.027)	(0.032)	(0.028)	(0.031)	(0.030)	(0.035)
Predetermined-year Control	✓	✓	✓	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1,896	1,260	1,847	1,861	1,679	1,833	1,763	1,596	1,693
R-squared	0.900	0.902	0.898	0.896	0.903	0.898	0.900	0.909	0.907

Notes: Table D7 reports results testing for subsample. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Column (1) reports the baseline results. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. Columns (2) to (9) respectively exclude the samples of provinces where the capital is not the economic center of the province, cities that are the best economically in their provinces but are not provincial capitals, separately planned cities, provincial sub-center cities, cities directly adjacent to Beijing and Shanghai, cities adjacent to the municipalities, cities adjacent to major regional centers, and ethnic autonomous regions.

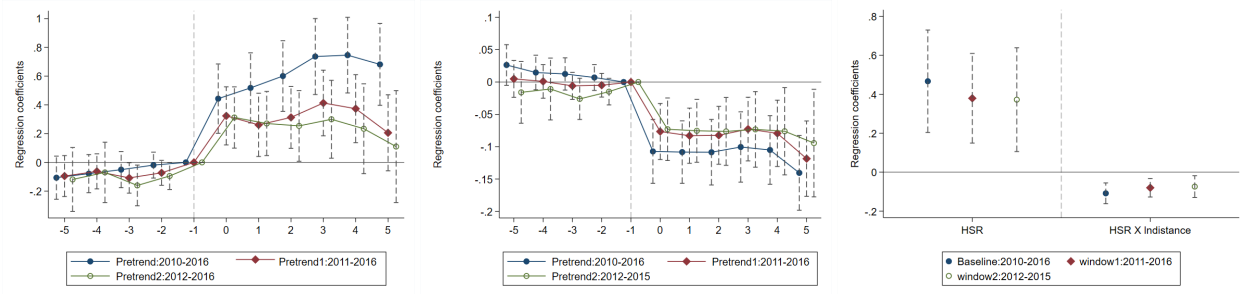
D.5 Alternative Model Specifications

In this section, we examine whether model specification would pose a threat our conclusions.

Firstly, as our protest information mainly comes from Weibo, one may be concerned that the data may be biased (particularly in the early period) due to the limited number of users. To address this concern, Figure D3 reports the baseline results and parallel trend test with different time windows. The results show that our conclusions are not affected by the choice of time window.

Secondly, we examine whether different clustering choices for standard errors would change our conclusion. We employ two-way clustering method to cluster standard errors at the prefecture-year, province, and province-year levels in Figure D4 and find similar results, for both the baseline and the parallel trend results.

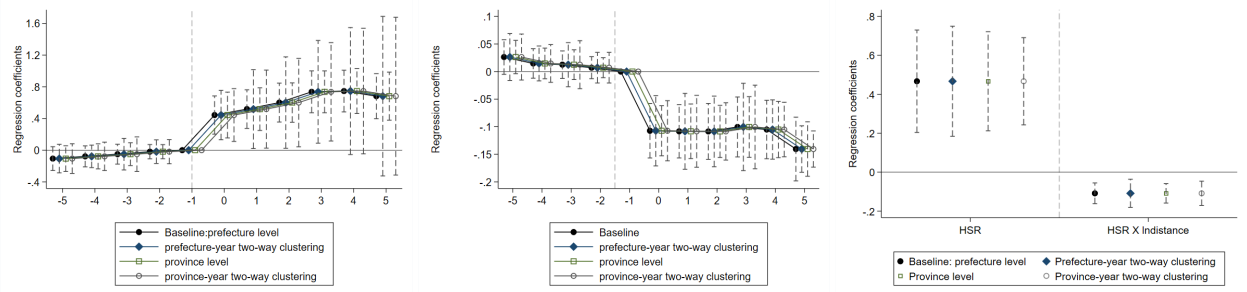
Thirdly, we examine whether different methods in the construction of dependent variables may affect our conclusion. In columns (2)-(5) of Table D8, we use the number of protests per year in each prefecture, the number of protests per million population, and the logarithm of the number of protests (plus one) as new dependent variables, respectively. This helps to rule out mechanical increase of protests due to rising population. We use a Poisson panel regression model in column (6). Furthermore, Table D9 presents results using alternative distance definitions. We find that all the results remain robust.



(A) **Pre-trend tests for the HSR term using alternative time windows.** (B) **Pre-trend tests for the interaction term using alternative time windows.** (C) **Baseline Results using alternative time windows.**

FIGURE D3 Testing for Alternative Time Windows.

Notes: These figures plot baseline results and pre-trend tests using alternative time windows. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). In Panel A-B, we plot pre-trend tests for the standardized number of protests using alternative time windows (2010-2016; 2011-2016; 2012-2015), while Panel C shows the baseline estimates using alternative time windows (2010-2016; 2011-2016; 2012-2015). We control for province-year, predetermined-year, and prefecture-level fixed effects for all specifications. The dotted line shows a 95% confidence interval.



(A) **Pre-trend tests for the HSR term using alternative clustering choices.**

(B) **Pre-trend tests for the interaction term using alternative clustering choices.**

(C) **Baseline estimates using alternative clustering choices**

FIGURE D4 Testing for Alternative Clustering Choices.

Notes: These figures plot the baseline results and pre-trend tests using four different clustering methods for standard errors, including prefecture-level clustering (used in the baseline), province-level clustering, prefecture-year two-way clustering, and province-year two-way clustering. The dependent variable is the standardized number of protests at the prefectural level (*Protestsd*). In Panel A-B, we plot pre-trend tests for the standardized number of protests using alternative clustering choices, while Panel C shows the baseline estimates using alternative clustering choices. We control for province-year, predetermined-year, and prefecture-level fixed effects for all specifications. The dotted line shows a 95% confidence interval.

TABLE D8 Robustness tests for alternative model specifications.

	OLS					Possion
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variables	*Protestsd*	*Protest*	*Protest per million people 1*	*Protest per million people 2*	*lnProtest*	*Protest*
HSR	0.467***	17.781***	4.364***	3.093***	0.267**	0.122*
	(0.134)	(5.089)	(1.124)	(1.019)	(0.103)	(0.068)
HSR×*lndistance*	-0.109***	-4.139***	-0.914***	-0.667***	-0.054**	-0.037***
	(0.027)	(1.031)	(0.227)	(0.209)	(0.022)	(0.014)
Predetermined-year Control	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓
Observations	1,896	1,896	1,896	1,896	1,896	1,887
R-squared	0.900	0.900	0.846	0.835	0.938	0.850

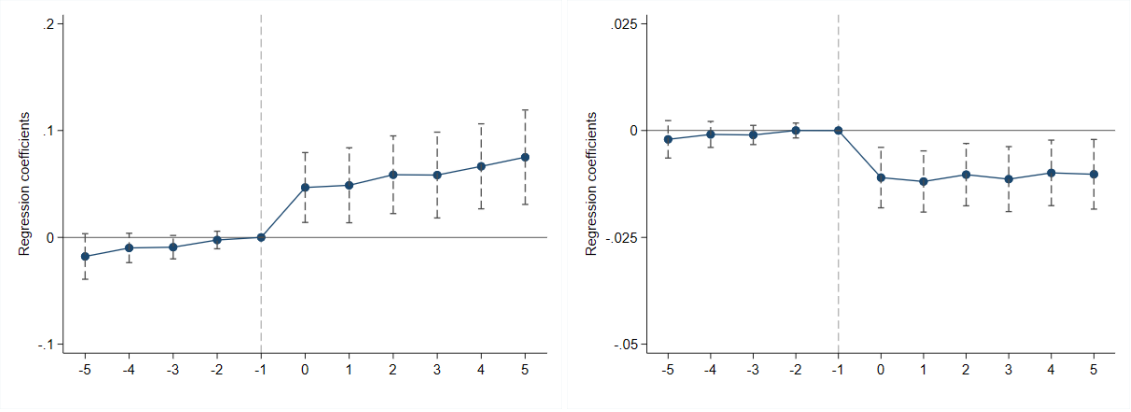
Notes: Table D8 reports results testing for alternative model specifications. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Columns (1)-(5) employ the OLS fixed effects model, while column (6) utilizes the panel Poisson model. Column (1) uses the standardized number of protests (*Protestsd*); column (2) uses the real number of protests (*Protest*); columns (3)-(4) divide the protest measure by the total population (in millions) (*Protest per million people*), where column (3) uses population data from the *Cities Statistical Yearbook*, while column (4) uses population data from *Global Gridded Population Dataset* (Liu et al., 2024); column (5) uses $lnProtest = ln(Protest + 1)$ as the outcome variable. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital.

TABLE D9 Robustness tests for alternative distance measurements.

	(1)	(2)	(3)	(4)
HSR	0.467***	0.178***	0.465***	0.131
	(0.134)	(0.068)	(0.133)	(0.088)
HSR×lndistance	-0.109***			
	(0.027)			
HSR×ln(0.001+distance)		-0.049***		
		(0.012)		
HSR×lndistance2			-0.109***	
			(0.027)	
HSR×Distance				-0.001**
				(0.001)
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,896	1,896	1,896	1,896
R-squared	0.900	0.900	0.900	0.899

Notes: Table D9 reports results testing for alternative distance measurements. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Dependent variables is the standardized number of protests (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The baseline results are presented in column (1), where variable *lndistance* is the logarithm of the distance from the prefecture to the provincial center plus one (units: kilometers). In column (2), variable *ln(0.001+distance)* is the logarithm of the distance from the prefecture to the provincial center, plus 0.001. Column (3) introduces variable *lndistance2*, which represents the logarithm of the distance from the prefecture to the provincial center, after defining the distance from the provincial capital city itself as 1. Finally, in column (4), variable *Distance* represents the distance from the prefecture to the provincial center (units: kilometers).

Appendix E More on Channels



(A) **Pre-trend tests for float ratio (*HSR* term).** (B) **Pre-trend tests for float ratio (Interaction term).**

FIGURE E1 Pre-trend tests for float ratio.

Notes: This figure plots pre-trend tests using prefecture-year level data. The dependent variable is float ratio at the prefecture level. Panel A tests for single *HSR* term. Panel B tests for the interaction term *HSR*×*lndistance* in equation (2); All panels control for Province-year, predetermined-year, and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.

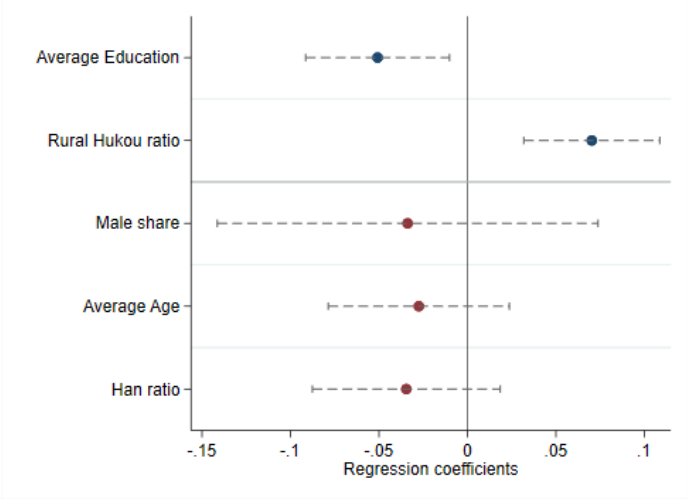
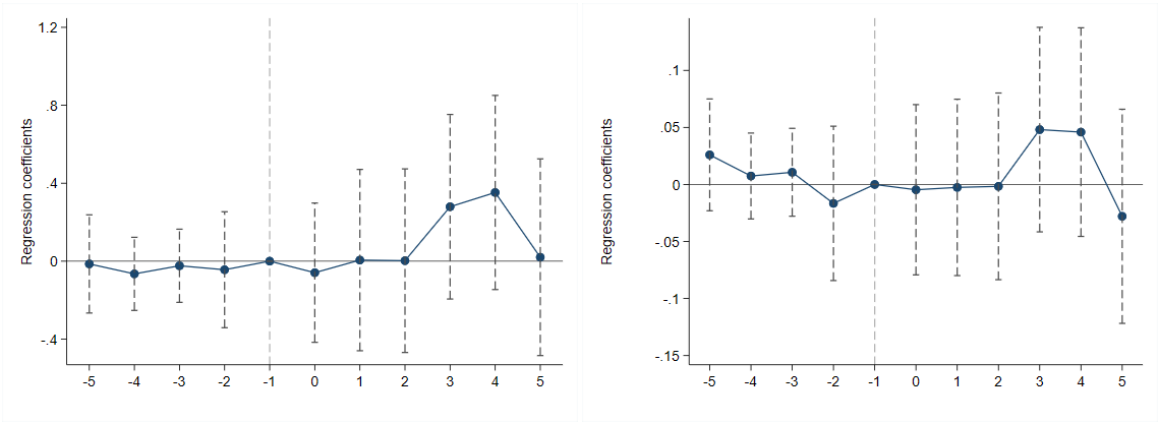


FIGURE E2 Relationship Between HSR Opening and the Average Characteristics of the Floating Population in Provincial Capitals.

Notes: This figure presents the relationship between the HSR opening and the average characteristics of the floating population in provincial capital cities using China Migrants Dynamic Survey (CMDS) 2011-2016 dataset. The dependent variables, listed from top to bottom, are the average years of education, the proportion of agricultural population, the male proportion, the average age, and the proportion of Han ethnicity in the floating population of provincial capitals over the years. All dependent variables are standardized. The independent variable is the number of in-province HSR stations connected to the provincial capital. Each point represents an estimated regression coefficient. The regression controls for fixed effects of provincial capitals and years. The dotted line indicates the 95% confidence interval.



(A) **Pre-trend tests for *HSR* term (*Petition*).** (B) **Pre-trend tests for interaction term (*Petition*).**

FIGURE E3 Pre-trend tests for the impact of HSR on the protests about upward petitions.

Notes: This figure plots pre-trend tests for the impact of the HSR on protests about upward petitions. Dependent variables consist of the standardized number of protests related to petitions (*Petition*). We obtain these kinds of protest events by adding extra protest classifications. Panel A test for single *HSR* term; while Panel B test for interaction term *HSR*×*lndistance*. All panels control for pre-determined characters, Province-year fixed effects, and prefecture fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.

TABLE E1 Falsifying monitoring effects

Dependent variable	(1) *Investigate official*	(2) *Weighted Investigate*	(3) *Protestsd*	(4)) *Protestsd*
HSR	-0.070	0.108	0.273**	0.278**
	(1.953)	(4.273)	(0.133)	(0.131)
	[2.258]	[5.097]	[0.114]	[0.113]
HSR×*lndistance*	-0.143	-0.328	-0.049*	-0.050*
	(0.427)	(0.915)	(0.028)	(0.027)
	[0.474]	[1.049]	[0.024]	[0.024]
HSR×*lndistance*×*Investigate*			-0.000	
			(0.000)	
			[0.001]	
HSR×*lndistance*×*Weighted Investigate*				-0.000
				(0.000)
				[0.000]
Controls	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,354	1,354	1,354	1,354
R-squared	0.818	0.797	0.940	0.940
Outcome mean	9.902	21.820	0.201	0.201

Notes: Table E1 reports the results testing the relation between HSR opening and investigation into local government officials. The data covers the period from 2012 to 2016. The dependent variable in column (1) is the number of corruption cases involving officials investigated in the prefecture in each year (*Investigate*). In column (2), the dependent variable is the index of corruption cases involving officials investigated in each year, weighted by the officials' ranks (*Weighted Investigate*). The dependent variables in columns (3)-(4) are the standardized number of protests at the prefectural level (*Protestsd*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital. Robust standard errors in parentheses are clustered at the prefecture levels. Standard errors in brackets are Conley standard errors robust for spatial correlation, assuming a cutoff window of 500 km and a serial correlation of 6 years. ***p< 0.01, **p <0.05, *p < 0.1.

TABLE E2 The heterogeneous impact of HSR openings on industrial structure and wages, housing prices.

	(1)	(2)	(3)	(4)	(5)
	The industry of			The logarithmic price of	
Dependent variable	*Primary*	*Secondary*	*Tertiary*	*Wage*	*House*
HSR	1.045	-1.259	0.231	-0.012	0.011
	(0.645)	(1.604)	(1.689)	(0.019)	(0.034)
HSR×*lndistance*	-0.218*	0.521	-0.308	0.004	-0.003
	(0.129)	(0.334)	(0.349)	(0.004)	(0.007)
Prefecture FE	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓
Observations	1,922	1,943	1,943	1,938	1,896
R-squared	0.948	0.934	0.928	0.969	0.918
Outcome mean	2.616	45.390	52.021	10.659	8.475

Notes: Table E2 reports the heterogeneous impact of HSR openings on industrial structure and wages, housing prices. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. In columns (1)-(3), the dependent variables represent the proportions of the primary, secondary, and tertiary industries in GDP. The dependent variables in columns (4)-(5) are the average wage and the logarithmic price of the house, respectively. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital.

TABLE E3 Alternative explanations: HSR-related protest.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Standard Number of Protest			Number of Protest			
Dependent variable	All	Non HSR	HSR	All	Non HSR	HSR	Protestsd
HSR	0.467***	0.632***	0.348	17.781***	16.884***	1.073	
	(0.134)	(0.117)	(0.216)	(5.089)	(3.129)	(0.666)	
HSR×*lndistance*	-0.109***	-0.138***	-0.073*	-4.139***	-3.680***	-0.226*	
	(0.027)	(0.026)	(0.044)	(1.031)	(0.685)	(0.136)	
Instant HSR							-0.419
							(0.306)
Instant HSR×*lndistance*							0.078
							(0.063)
Observations	1,896	1,896	1,896	1,896	1,896	1,896	1,896
R-squared	0.900	0.896	0.393	0.900	0.896	0.393	0.900
Control	✓	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓	✓

Notes: Table E3 tests the alternative explanations for HSR-related protest.***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels.

TABLE E4 Alternative explanations: financial welfare improvements after HSR opening.

Dependent variable	(1) Social care	(2) Internet	(3) Park	(4) Book	(5) Bus	(6) Education finance (%)	(7) Science finance (%)
HSR	1.504	-0.049	-99.39	-7.973	2.862	-0.233	4.011
	(2.379)	(0.050)	(186.459)	(17.206)	(2.452)	(0.470)	(3.684)
HSR×*lndistance*	-0.429	0.008	-1.090	0.933	-0.431	0.010	-0.566
	(0.519)	(0.010)	(21.190)	(3.540)	(0.387)	(0.095)	(0.734)
Prefecture FE	✓	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓	✓
Observations	1,936	1,929	1,931	1,664	1,663	1,942	1,943
R-squared	0.779	0.740	0.934	0.535	0.671	0.544	0.897
Outcome mean	42.615	0.121	195.347	9.929	9.078	2.907	38.416

Notes: Table E4 tests the alternative explanations for financial welfare improvements after HSR opening. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. The dependent variables in columns (1)-(7) are the number of hospital beds per capita, the proportion of Internet access users, the per capita park area (unit: square kilometers), the per capita number of public books, the per capita number of buses, the proportion of education expenditure in fiscal revenue, and the proportion of science and technology expenditure in fiscal revenue, all of which refer to per 10,000 people. The data primarily come from the *Chinese City Statistical Yearbook*. *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital.