State Infrastructure and Political Geography of Protest [*]

Siyuan Fan [†] Wenyi Lu [‡]

Abstract

How does state infrastructure shape social unrest? Leveraging the staggered rollout of high-speed rail (HSR) in China, we find that HSR reduces protests in peripheral areas while causing more protests in provincial centers. This spatially heterogeneous effect is primarily driven by protests involving migrant workers. We then develop a theoretical framework outlining four potential channels. Empirical evidence from both administrative and survey data predominantly supports the agglomeration effect, where HSR facilitates migration to central areas and triggers protests. We also provide suggestive evidence for the deterrence effect, whereby improved connectivity enhances state presence and suppresses protests in remote regions. Our findings highlight the political implication of state infrastructure: while intended to enhance economic integration and state control, it may relocate dissent to politically sensitive urban centers, posing new challenges for authoritarian rule.

Keywords: State infrastructure; Protest; Transportation; Migration; Political geography.

JEL Codes: D74, O18, P25, R12, R40

[*]We are thankful to Lucy Barnes, Leah Platt Boustan, Yiming Cao, Ting Chen, Donald R. Davis, Melissa Dell, Quoc-Anh Do, Dave Donaldson, Lukas Hensel, Junyan Jiang, Yu Qin, Mark Rosenzweig, Awa Ambra Seck, Romain Wacziarg, Yuhua Wang, Tianyang Xi, Yiqing Xu, and Yang You for helpful suggestions. We benefit from discussing with Romain Angotti, Clement de Chaisemartin, and Peter Hull concerning econometric specification. We also thank seminar/conference participants at RES Annual Conference 2024, Asian Meeting of the Econometric Society (AMES2024), Renmin University of China, and Peking University for helpful discussions. All remaining errors are our own.

[†]UC San Diego, 9500 Gilman Drive, La Jolla, CA; email: sif004@ucsd.edu.

[‡]New York University, 70 Washington Sq S, New York, NY; email: wenyilu@nyu.edu.

1 Introduction

Economic modernization is often accompanied by shifts in political geography (Polanyi et al., 1944; Huntington, 2006). State infrastructure, including highways, railways, post offices, and so on, acts as an important tool for the state to enhance economic ties and political control over its territory, particularly remote areas (Mann, 1984). The modernization of state infrastructure has the potential to reshape regional economic and political dynamics profoundly. While some scholars highlight the role of state infrastructure, such as railways, in promoting trade and facilitating economic development (Faber, 2014; Donaldson, 2018; Brancaccio et al., 2020; Allen and Arkolakis, 2022), others emphasize its function in extending state control through monitoring and repression, as evidenced by roads, probes, and police stations (Xu, 2021; Chang and Wang, 2024; González et al., 2025). These functions are encapsulated in the "state presence" theory, emphasizing the state's incessant efforts to build infrastructure so as to penetrate society, extract rents, exerting social control, and foster citizens' compliance (Mann, 1984; Shue, 1990).

However, there remains no clear consensus on how state 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). This article offers an alternative perspective that state infrastructure can reshape the *political geography* of protest: suppressing protests in remote regions while inciting them in the center.

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. Seen in this light, our study reveals a paradoxical result: while being intended to enhance economic linkage and state power, state infrastructure may backfire, inducing more protests near the capital and disturbing authoritarian rule.

China's High-Speed Rail (HSR) program acts as an ideal case to examine the question. 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.

To analyze its political implications, 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. We manually collected data from the China Railway Yearbook to determine HSR access in each prefecture in any year between 2010 and 2016. Protest events are quantified using the Collective Action from Social Media (CASM) dataset, aggregated to the prefecture-year level (Zhang and Pan, 2019). We also employ the China Labor Bulletin (CLB) dataset as further verification. As HSR extends from provincial capitals to local prefectures, we focus on within-province variation and examine how HSR affects protest incidence in regions with varying distances to provincial capitals.

Using difference-in-differences (DID) as our empirical strategy, we explore how HSR opening can reshape the incidence of protests across regions within each province. The 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.

We further relax the linearity assumption by analyzing the respective effects of HSR opening on different distance sections, and find strong positive effect for central regions while more negative effect for distant areas. Using heterogeneity-robust estimators that address concerns about treatment effects varying over time and/or across cohorts (e.g., De Chaisemartin and d'Haultfoeuille (2024)), we continue to observe consistent results. Monthly protest data further corroborates these results. We also categorize protests by motive, finding they are predominantly driven by issues on wages, housing, education, and medicine, suggesting significant involvement of migrant workers. Verification using the CLB dataset supports the spatial heterogeneity of HSR's effects on strikes. Moreover, such strikes only emerge in pro-migrant industries following HSR opening, which again implies the significant role of migrants.

Our results are robust to a battery of placebo tests and robustness checks. We complement the binary indicator for HSR opening with a continuous measure: the associated growth in market access (MA) from HSR expansion. Increasing MA give rise to protests in central regions while reduce resistance in peripheral areas, which again corroborate our baseline results. To account for potential non-random expectation in HSR opening and MA, we employ the recentered-IV ap-

proach by Borusyak and Hull (2023), and find consistent results. We then exclude the influence of other confounding factors, including political, fiscal, technological, and geographical variations across prefecture-year units. In addition, we discuss a few cases where the political center (provincial capital) may not hold sufficient economic centripetal force and obtain robust findings, highlighting the interplay between economic and political forces.

To explore the sources of HSR's spatially heterogeneous effects, we develop a theoretical framework of transportation networks and protests, extending the work of Passarelli and Tabellini (2017). In this framework, HSR opening can affect protests through four main channels. First, the *agglomeration effect* posits that more migrants flow towards the center, resulting in a concentration of population and economic activities that may escalate conflicts. Conversely, the *spillover effect* suggests that peripheral areas benefit economically from HSR at the cost of the center. Third, the *accountability effect* argues that stronger oversight of local officials may reduce protests through improved governance. Finally, the *deterrence effect* considers heightened state presence to deter protest participation. Empirically, we provide solid evidence for the *agglomeration effect* that HSR attracts capital and labor to central areas, giving rise to economic and population agglomeration. A higher concentration of migrants, mainly low-skilled and from rural areas, is associated with increased protest events, particularly in regions with severe discrimination. The above findings directly reject the *spillover effect*. We also rule out the *accountability effect*, finding no evidence that HSR enhances local officials' accountability or improves governance in peripheral regions. Instead, we provide suggested evidence for the *deterrence effect* that HSR opening is associated with lower political trust, who also become less likely to express their political views. Furthermore, we exclude alternative explanations, including HSR-related protests, commuting for protests, or changes in public expenditure.

This study fits directly into the literature on the implications of transportation infrastructure. Since Fogel (1964)'s seminal analysis of the social savings from railroads in the United States, subsequent research has extensively examined the economic impacts of transportation infrastructure (like Duranton, Morrow, and Turner (2014), see Redding and Turner (2015) for a literature review). Some studies focus on the overflow of resources and industrial activities (See e.g., Charnoz, Lelarge, and Trevien (2018); Lin (2017); Baum-Snow et al. (2017)), while others emphasize the agglomeration of human capital (See e.g., Faber (2014); Dong (2018); Dong, Zheng, and Kahn (2020)), yet discussions on its political ramifications remain limited. A few studies have investigated how infrastructure influences election outcomes (Voigtländer and Voth, 2022; Nall, 2015; Akbulut-Yuksel, Okoye, and Turan, 2024), crimes (Calamunci and Lonsky, 2024), and state repression (González et al., 2025), but little is known about political uprisings, a critical dimension for state stability.[1] This study fills the gap by examining HSR openings and identifying the

1. Cao and Chen (2022) find rising social conflicts during canal abolishment in ancient China. We differentiate from their work by focusing on the interaction between economic centripetal force and political repressive power, and the resulting spatially heterogeneous outcome.

mechanisms through which they may mitigate or exacerbate protests.

Moreover, our study helps clarify a few controversies on the benefits of transportation infrastructure. While some scholars contend that such infrastructure can spur economic growth in peripheral regions (Zheng and Kahn, 2013; Lin, 2017; Ahlfeldt and Feddersen, 2018; Bernard, Moxnes, and Saito, 2019), others are less convinced and find ambiguous results (Faber, 2014; Asher and Novosad, 2020; Banerjee, Duflo, and Qian, 2020; Borusyak and Hull, 2023). Qin (2017) even find a disproportionate benefit for large cities. We highlight geographical distance as a critical determinant of heterogeneity, similar to Brinkman and Lin (2024), but offer a more comprehensive framework for the underlying patterns.

Our study builds upon the work of Campante, Do, and Guimaraes (2019), which highlights the role of spatial distribution in protests and revolutions: in non-democracies, citizens are more likely to protest in capitals due to their greater political influence. Our research diverges from theirs in three important aspects. First, they focus on natural distances to capital cities and examine capital changes as variations. In contrast, we consider the role of state infrastructure (transportation networks) in changing effective travel distance from various places to the capital. Second, while their work help explain the current protest distribution (more protests in the center), we aim to explore its dynamics, showing how economic modernization would reshape its spatial distribution. Moreover, we offer a variety of mechanisms through which state infrastructure could reshape protest dynamics, which helps understand the different forces at play and illustrate its future evolution. Third, we integrate the mechanisms with a model that combines the spatial framework with rational protest participation (Passarelli and Tabellini, 2017). This helps provide solid micro-level foundations for changes in protest geography.

Our work also draws on historical studies that highlight the importance of transportation in shaping protests and revolts. For instance, Traugott (2010) examines the impact of transportation infrastructure on 19th-century French protests, attributing reduced protests after 1871 to the extension of railroads. Our study differs in two features. First, they emphasize the role of transportation in alleviating food shortages and promoting market integration, yet we focus on how it affects population mobility and state presence. This also accounts for our different findings: while they find unanimous decrease in protests, we show a spatially heterogeneous outcome (rising in the center while decreasing in the periphery). Second, while they didn't offer any empirical support for the story, we leverage a comprehensive dataset to provide detailed empirical evidence on how state infrastructure influences the change in political geography in the context of a modern authoritarian state.

Additionally, we contribute to the literature on migration by examining the political implications of intra-regional migration, a topic often overlooked in existing research.[2] HSR-induced

2. Ample evidence has been provided on large-scale, long-term, trans-regional immigration, like its impact on regime change (Karadja and Prawitz, 2019), voting behavior (Barone et al., 2016), urban disorder (Castells-Quintana et al., 2022),

migration in China provides a unique opportunity to examine these dynamics and their complex channels. Focusing on intra-regional migration in China, current research mainly explores its economic impacts, such as economic development (Tombe and Zhu, 2019), firm productivity (Imbert et al., 2022), and wage disparities (An et al., 2024). We extend the literature by examining how migration reshapes the political geography of protests.

We also complement the research on the source of protests in authoritarian states. Sociopolitical literature highlights the role of participatory willingness (driven by internal grievances) and protest capability (shaped by external conditions) in triggering protests respectively (Gurr, 1970; McCarthy and Zald, 1977; McAdam, 1999).[3] Recent studies have introduced technological progress as external shocks to examine their impact on protests, like communication technologies (Enikolopov, Makarin, and Petrova, 2020), mobile phones (Manacorda and Tesei, 2020), and general social media (Qin, Strömberg, and Wu, 2024). Our contribution differs from these works by exploring how state infrastructure influences protests, distinguishing the contribution of participatory willingness and protest capability, thereby linking the two strands of literature together.

More broadly, this study links with the literature on state presence that focuses mainly on how the state projects power across its 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, however, have documented its potential negative effects on social stability (Ying, 2021; Chiovelli et al., 2024; Calamunci and Lonsky, 2024). This study contributes to this emerging discourse by examining the unintended consequences of state presence: while facilitating state control in peripheral regions, HSR inadvertently redirects protests to the center, resulting in heterogeneous outcomes. While some scholars have emphasized heterogeneity in political outcome, this is among the first research to highlight the distributional impacts of state presence on protest dynamics.[4]

The rest of this paper is laid out as follows. Section 2 presents a theoretical framework and outlines our main empirical hypotheses. Section 3 introduces the Chinese HSR system and its politico-economic context. Section 4 details the data and outlines our empirical strategy. Section 5 examines how HSR opening exerts a heterogeneous impact on protests among different regions. In section 6, we offer detailed empirical evidence to verify the channels and discuss alternative explanations. Section 7 concludes.

and even the civil rights movement (Calderon et al., 2023).

3. Gurr (1970) focuses on protest participatory willingness, namely "relative deprivation", followed by empirical studies like Cantoni et al. (2019) and Bursztyn et al. (2021). The political process theory (McCarthy and Zald, 1977; McAdam, 1999), however, emphasizes protest capability, particularly the state's deterrence power.

4. Bautista et al. (2023) examines heterogeneous state repression in Chile, focusing on locations of military bases. Another related study is Schuler and Truong (2020) that discussed the differential effect of the Internet on protests, due mainly to rural-urban heterogeneity.

2 Theoretical Framework

State infrastructure can affect protest participation both positively and negatively through different forces. Building on the work of Passarelli and Tabellini (2017), we present a simple theoretical framework in which transportation infrastructure affects protest participation by fostering migration (*agglomeration effect*), facilitating resource spillover (*spillover effect*), monitoring local officials and improving governance (*accountability effect*), and deterring local residents (*deterrence effect*). Finally, we are able to show how different effects jointly affect protest participation and reshape the spatial distribution of protests. Overall, this framework provides useful micro-level foundations for our empirical analysis and brings about several insightful predictions that allow us to disentangle the mechanisms. We present a concise exposition of the framework below. Please see Appendix A for the full model setup, derivations, and other details.

Consider an economy with one central city (e.g., a provincial capital) and a large number of smaller peripheral cities, with a continuum of citizens uniformly distributed on the line. Citizens rationally decide whether to work in their hometown or the central city, trading off the income gap and transportation costs. After making this workplace decision, they unilaterally decide whether to participate in protests, weighing the costs (including time, resources, and the risk of arrest) against the benefits. The costs are primarily determined by external conditions, while the benefits increase with both the level of grievance and the number of other participants (i.e., due to complementarity). Thus, the protest participation rate is determined by the fixed point of the equilibrium condition. Comparative statics show that protest participation increases with (i) a larger pool of potential participants, (ii) higher levels of grievance, and (iii) lower repression. Based on this framework, we can analyze how HSR opening reshapes the spatial distribution of protests.

First, the expansion of HSR network can enhance connectivity between the center and other prefectures. With lower transportation cost, more citizens will leave their hometown and move to the central city. As a result, there would be more potential protest participants in the central city while less in remote areas.

Hypothesis 1 (Agglomeration Effect) *HSR opening can foster migration, thus*

- *increasing protests in central areas by attracting more migrants;*
- *having little (ambiguous) influence for middle regions;*
- *reducing protests in remote areas by pushing residents away.*

Meanwhile, HSR opening may readjust the wages and housing prices in different regions to balance the migration flow: in central cities, rising labor supply and housing demand may drive up housing costs while reducing wages; in peripheral areas, wages might increase while housing

costs might decline. Consequently, people in the center find themselves worse-off and protest more, whereas people in the periphery live a better life and become less willing to protest.

Hypothesis 2 (Spillover Effect) *HSR opening might 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.*

These two hypotheses present opposite explanations for the same finding. For Hypothesis 1 to hold, we would observe rising agglomeration in the center, while Hypothesis 2 predicts resource spillover to the periphery. Meanwhile, Hypothesis 1 ascribes rising protests in the center to more migrants, while Hypothesis 2 points to worse living conditions. These predictions are crucial for us to partial out which channel is at play in our empirical analysis.

For remote areas, the extension of HSR manifests rising state presence. On one hand, its extension facilitates central inspection into local officials. Expecting a higher likelihood of being detected, local officials would rationally reduce corruption and improve governance. Thus, HSR may enhance accountability and palliate resistance in remote areas.

Hypothesis 3 (Accountability Effect) *HSR opening can facilitate investigations into local officials, alleviating corruption and reducing protests in remote areas.*

On the other hand, its extension also marks rising coercive state power, facilitating repression against resistance. Expecting a higher likelihood of being caught, individuals would rationally reduce protest participation.

Hypothesis 4 (Deterrence Effect) *HSR opening can extend state presence, deterring residents from launching protests, thus reducing protests in remote areas.*

These two hypotheses focus on the political meaning of HSR opening, which also offer different predictions. Hypothesis 3 offers a positive blueprint, attributing reduced protests in the periphery to improved governance (lower levels of grievance). On the contrary, Hypothesis 4 presents a much more pessimistic view, arguing that people protest less as they are deterred by rising repression force.

To sum up, the four hypotheses shed light on different meanings of HSR extension, yet they are consistent that there would be rising protests in the center but less in the periphery following HSR opening, shown in Hypothesis 5. In section 5, we will provide solid empirical evidence for this common prediction, and then investigate the extent to which each hypothesis can account for this finding in section 6.

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

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).[5] 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).[6]

China's investment in HSR has grown exponentially, soaring from $14 billion in 2004 to a substantial $719 billion by 2022, taking up a significant portion of the government's budget.[7] Concerns were raised about its safety, occupancy rates, and environmental impacts. Nevertheless, HSR remains a favored mode of transport due to its speed, reliability, and comfort, delivering a 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,

5. 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.

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

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

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.[8] 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 large territory and household registration system (*hukou*) impede inter-provincial migration.[9] 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[10] 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.[11] 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.

8. This assumption does not restrict our main finding, as shown in Table D9 in the Online Appendix.

9. 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.

10. 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).

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

Local governments, however, possess minimal decision-making authority during the route-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.[12]

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.

3.3 Rural-urban Income Gap and Discrimination

HSR opening could afford much convenience to the rural-urban migration flow. While enabling economic opportunities, it also underscores systemic disparities between urban and rural areas. In 2010, the average annual income for urban residents is 19,109 RMB, while that for rural residents is only 5,919 RMB, less than one third of their urban counterparts ($\frac{5919}{19019} \approx 0.31$). Such rural-urban income disparity, together with huge gap in culture and consequent differences in values, life styles, and dialects, have resulted in persistent discrimination against migrant workers in Chinese central cities. Such discrimination, coupled with limited labor protections, and inequitable wages, has hindered the integration of migrant workers into urban society (Rho, 2023), motivating strong grievances. The migration, initially underpinned by hope, culminates in greater disillusionment and exacerbates urban protests. See Appendix B for details.

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. To ensure the robustness of our

12. 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.

findings, we extend our analysis to prefecture-month-level datasets.

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.

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.

The CASM dataset has several advantages. First, it has high credibility. Zhang and Pan (2019) evaluate its performance through cross-validation, out-of-sample validation, and comparisons with other protest datasets, such as the Global Database of Events, Language, and Tone (GDELT). They show the main features of the CASM dataset align with all above, and more importantly, CASM offers broader coverage, documenting a greater number of daily protests that are ubiquitous in the daily lives of Chinese citizens.[13] Second, it is minimally influenced by censorship. Unlike state media that is heavily censored in China, Weibo's censorship was not stringent until 2017. From 2010 to 2016, many sensitive, provocative, and anti-government posts appeared on Weibo without being censored.[14] Moreover, post hoc content removal does not substantially influence the data (Zhang and Pan, 2019). Third, the dataset provides extensive details. The CASM categorizes protest events into 11 types based on motives and 3 primary forms based on action types. Figure B6 shows the structure of different kinds of protests and their temporal variation.

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 predict social media penetration, so even if there is potential regional biases in media usage, it cannot confound our estimation. Second, we use alternative data sources to complement CASM.

13. GDELT puts more coverage on ethnic conflicts (40%), which is more plausible in analyzing anti-regime protests and repression, like Beraja, Kao, Yang, and Yuchtman (2023). In comparison, CASM concentrates more on daily protests like unpaid-wage (29%), property (26%), and land-related (17%) protests, which fits our focus.

14. Media censorship has expanded since 2017 when President *Xi* established himself as the "core" of CPC and enhanced social control. For instance, the Cybersecurity Law in 2017 induced heavier regulation over the Internet and a stronger ban on using VPNs (Chen and Yang, 2019). Hence, media reports on protest events became rare after 2017.

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).[15] Table B2 presents descriptive statistics for our main variables in the CLB dataset. As shown in Figure B7, 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

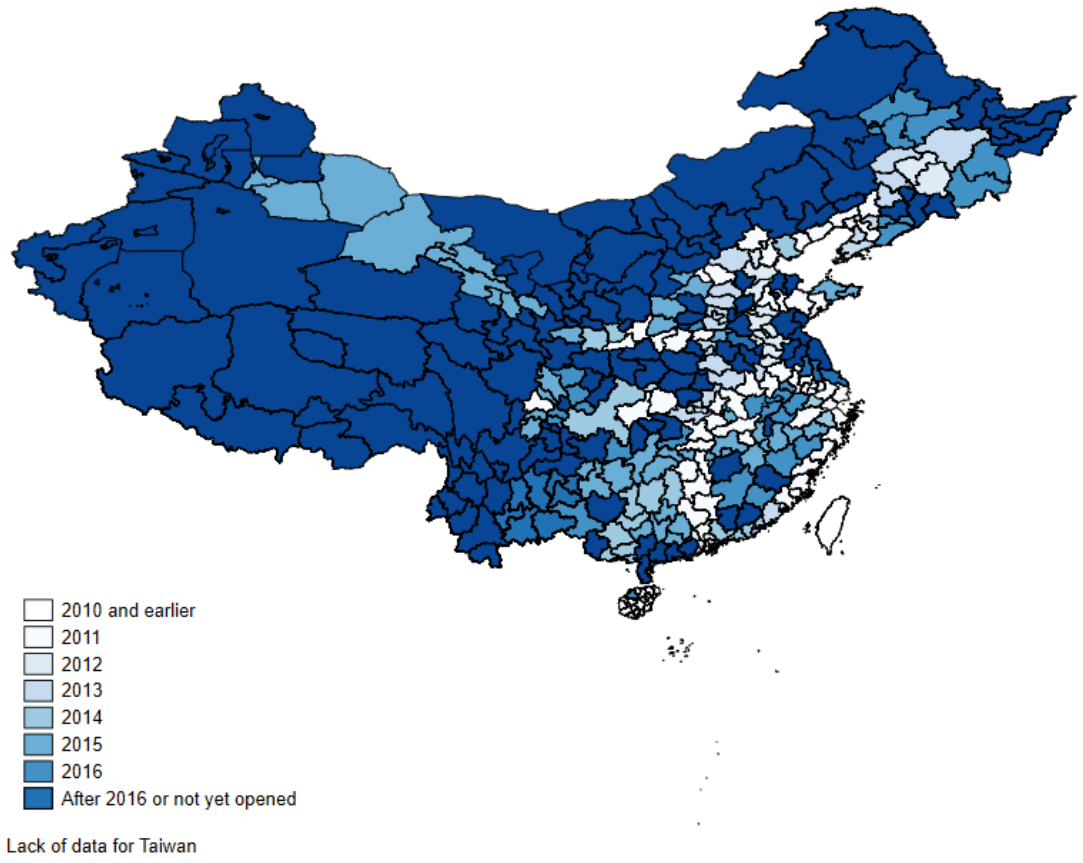
We obtain the data on whether the prefecture has access to HSR in a certain year from the China Railway Yearbook, encompassing details such as construction commencement dates, opening dates, route lengths, designated speeds, and ridership. We cross-verify the stoppage points along each operational line 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

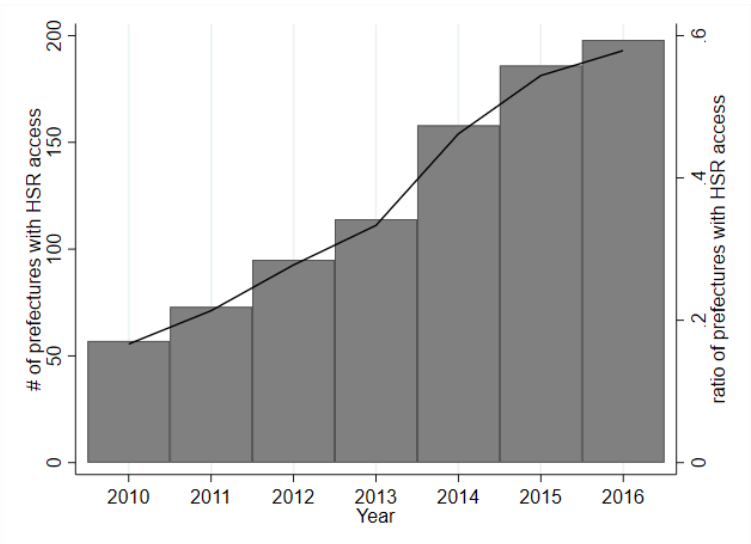
China City-specific Characteristics. To control for regional influences, we incorporate predetermined prefectural characteristics from the *China City Statistical Yearbook*, including the proportion of the primary sector, that of the secondary sector, per capita GDP, population, and urban area in 2010. 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, capturing a wide range of aspects of Chinese citizens' lives. We use it 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

15. 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.

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. Lastly, we leverage the Chinese Social Survey (CSS) to measure perceived discrimination of migrant workers. See Appendix B for details of the survey data.

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. We plot its spatial distribution in Figure B8, showing that most protests occurred in economically developed regions and populous provinces, such as Sichuan, Shaanxi, Henan, and Jiangsu.

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: Prefecture-month Outcomes					
Protest_month	22,752	2.399	3.719	0	50
Protestsd_month	22,752	0	1	-0.645	12.801
Panel D: 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

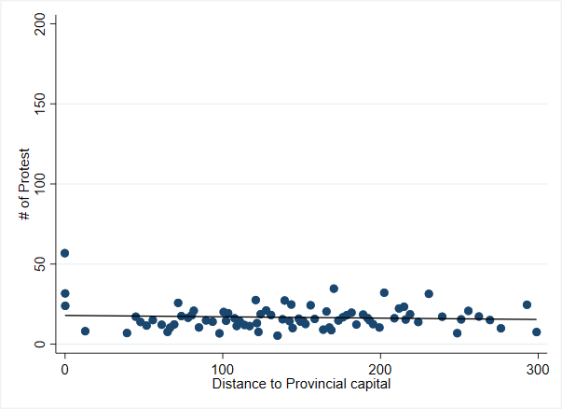
Notes: Table 1 presents descriptive statistics for our main variables.

[a]: $Protest$ is the number of protests, while $Protestsd$ is the standardized number.

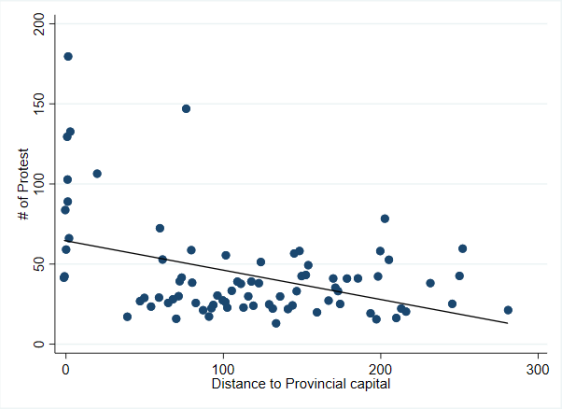
[b]: We define $Area$ as the logarithm of the prefecture's land area (unit: square kilometers).

Based on the raw data, we compare the geography of protests before and after HSR opening in Figure 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. Figure B8 also confirms this pattern: within each

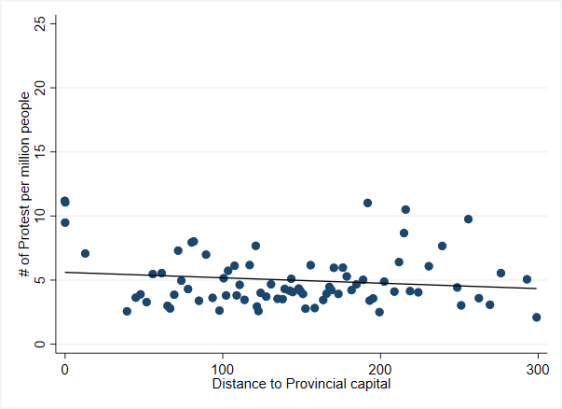
province, the provincial capital has substantially more protests than other prefectures in 2016. For causal identification, we employ the Difference-in-Differences (DID) method in subsequent analyses.



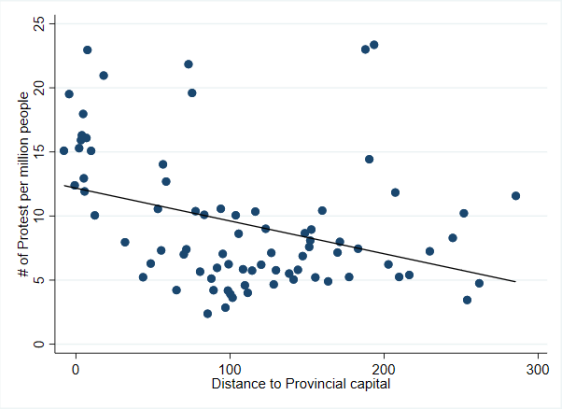
(A) **The geographical distribution of protest events before HSR opening.**



(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.

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:

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

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 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. We also employ alternative clustering choices as robustness checks.

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 Equation 1 and Equation 2 is the inclusion of the interaction term between distance to the provincial capital ($lndistance_c$) and HSR. 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. The results are robust with different measures like logged distance (without plus 1) or actual distance. 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.

Identification Assumptions Our identification does not require the exogeneity of HSR opening, though this assumption likely holds at the prefecture level. 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.

Diff-in-diff estimation with staggered treatment can suffer from a bias when there are heterogeneous treatment effects over time or across subgroups. Since the two-way fixed effects (TWFE) estimator is a weighted sum of the average treatment effects (ATEs) in each group and time period, some ATEs may be given negative weights, causing the estimator to have a different sign than all the ATEs (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 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.

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 de-

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. Nonetheless, our findings are robust to other estimators like CSDID.

viations (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).

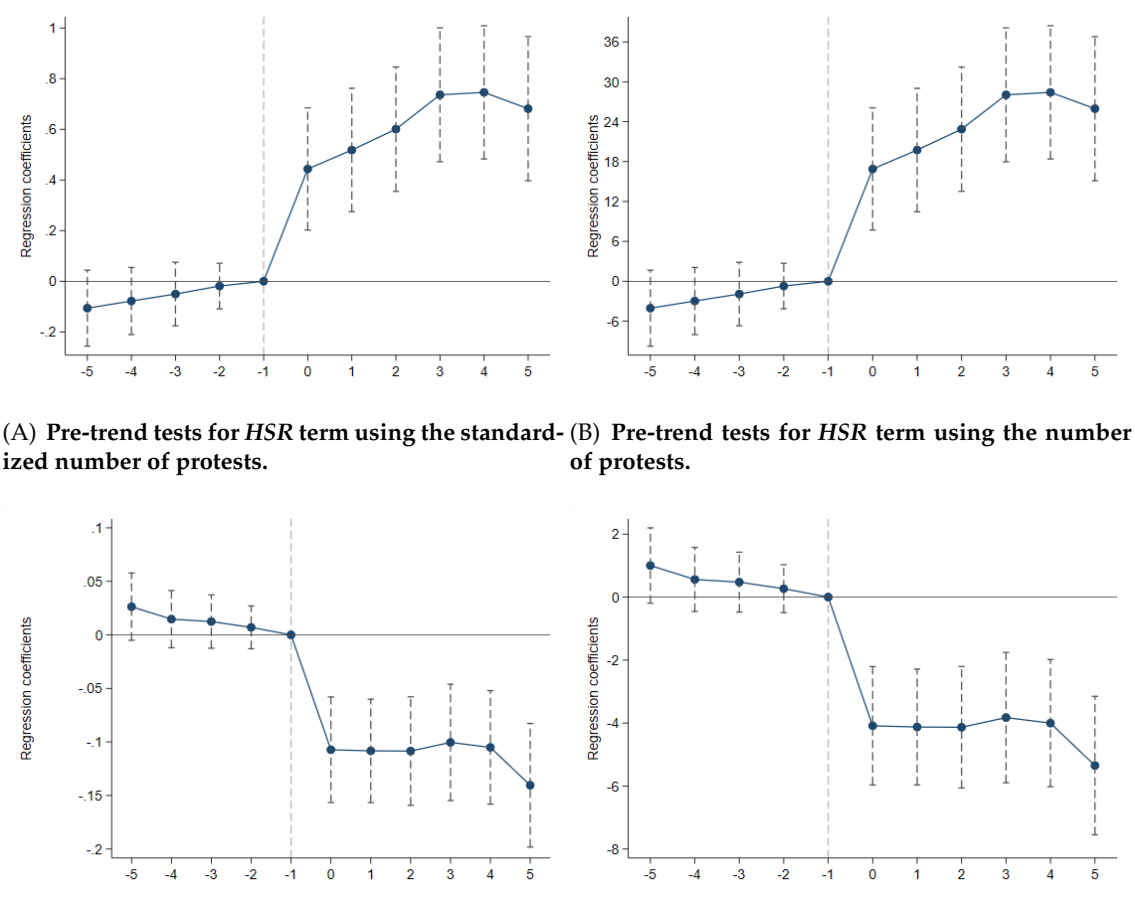
TABLE 2 Baseline Results

Dependent variable	*# of Protest (std)*						Protest per million people
	(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 per million people*). *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. 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.

We then compare our results with related works. Chang and Wang (2024) also use the CASM dataset, finding 1.384 fewer protests after the establishment of the Committee for Comprehensive Management of Public Security (CMPS) at the county level. Their result is equivalent to 0.269 s.d., smaller than our effect. Considering a prefecture has 8.5 counties on average, their result can also be regarded as 11.78 fewer protests for each CMPS. Therefore, the effect of HSR is at least as large as the CMPS program. Similarly, Manacorda and Tesei (2020) find a 4 percent fall in GDP growth leads to 2.3 more protests (per million people) between areas with mobile phone coverage and those without. Column (7) of Table 2 shows 4.364 more protests after HSR opening

in central regions and 0.914 fewer with distance doubled. Thus, the effect of HSR opening is roughly equivalent to a 7.59 percent decrease in GDP ($4.364/2.3 \times 4 = 7.59$), and doubled distance functions similarly as a 1.59 percent rise in GDP ($0.091/0.23 \times 4 = 1.59$) in alleviating protests. The result also demonstrates that our findings are not solely driven by mechanical population growth.



(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.

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

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

5.2 Using Monthly Data

To delve into the detailed temporal evolution of HSR's effect, we obtain monthly data on protests and estimate Equation 2. Table C3 shows a similar effect with our baseline. As we include stricter fixed effects, HSR's effect becomes more significant, both statistically and economically. In column (5), we control for the province-year-month and prefecture-month fixed effect along with the interaction terms between predetermined variables and year.[19] We find that, protests in central regions increased by 0.325 s.d. after HSR opening, but decreased gradually as distance doubled (by 0.074 s.d.). The result is equal to 1.209 (0.325×3.719) additional protests in the center each month (decreasing by 0.274). We plot the pre-trend tests using monthly data in Figure C2. Similarly, the two groups show minor differences in protest pre-trends, with protests rising sharply in central regions and declining with the geographical distance.

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.

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)).

19. The "month" fixed effect keeps within-month variations (not within the same year). Hence, we adopt the "year-month" fixed effect to keep variations within the same month of the same year. Likewise, the Province-year-month fixed effect compares samples within the same prefecture and the same year-month.

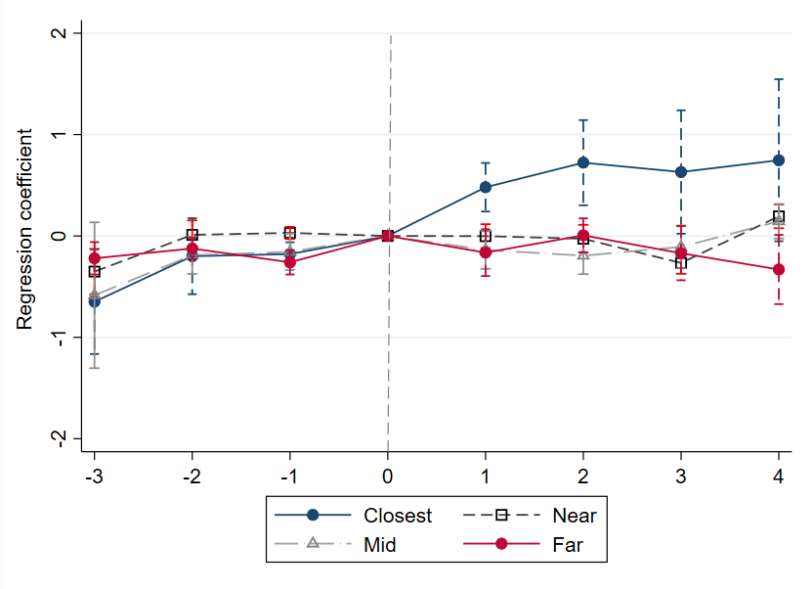


FIGURE 4 Event study in different groups with different distance to provincial capital.

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) over various years since the opening of high-speed rail 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.

5.3 Further Classification

Protests vary both in their causes (content) and intensity (form). People engage in protests for different reasons, ranging from political issues, like the Gwangju Uprising, the Prague Spring, the Arab Spring..., to social and economic problems like the "Occupy Wall Street" movement. Different kinds of protests carry distinct meanings, prompting the government to make distinct responses (Cantoni et al., 2024). 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 C7 and Figure C8. 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.

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.

In addition, the intensity of protests also differs. While some protests are moderate, others are more extreme, involving physical confrontations, even armed struggle, bloodshed, and mass injuries. We examine whether residents engage in drastic or moderate protests after HSR opening. 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), following Zhang and Pan (2019). 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 C4. All three forms of protest show similar trends, but disruptive ones are affected less by HSR opening. The pre-trend tests in Figure C3 and Figure C4 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.[20]

5.4 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 C5. 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. Similar results can be obtained using monthly-level data, as shown in Figure C6.

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

20. Previous studies have shown that politically related protests, especially coups, are more likely to be well-organized and long-planned (Cantoni et al., 2024). Economic and social issues, on the contrary, usually motivate people to launch protests in a sudden and end them soon.

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.

TABLE 4 Baseline Results: CLB dataset (Yearly data)

	(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.

5.5 Placebo Tests

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 C5. 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 C6. 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 C9 indicate that our baseline results are not significantly affected by this issue.

5.6 Robustness Checks

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

Long-difference, market-access, and recentered-IV. To alleviate biases arising from transitory shocks or serial correlation over time (Griliches and Hausman, 1986), we construct a cross-sectional dataset to examine the relationship between changes in HSR access and changes in protest events during 2010-2016. Columns (1) and (2) of Table D1 show consistent results with our baseline findings: changes in HSR opening have brought about significantly more protests in the provincial capital, while the effect wanes as the region moves farther from the capital.

To account for the "quality" of connectivity, we introduce a continuous measure of HSR exposure based on market access (MA) growth following Borusyak and Hull (2023). This approach captures the difference in travel speed of HSR lines and the number of cities they connect, with the assumption that HSR lines are randomly built conditional on the plans and line lengths. In column (3) of Table D1, we use $MA\ growth$ as an alternative measure of exposure to HSR and obtain similar findings (See Appendix D for details of the construction process).

To deal with endogeneity concerns with MA, we employ a recentered-IV approach for causal identification. Following Borusyak and Hull (2023), we recenter our estimates by controlling for the expected probability of HSR opening ($Expected\ MA\ growth$). Specifically, we generate a counterfactual network by permuting the completion status of built and unbuilt HSR lines in 2016, repeating this process 1,000 times to compute expected MA growth as the average across permutations. Using the difference between actual and expected MA growth as an instrument allows us to isolate exogenous variation and identify the causal effect of HSR on protest activity. As shown in columns (5) and (6), improved market access leads to an increase in protests in central regions, while reducing protests in peripheral areas. Alternatively, column (4) reports the effect of recentered MA growth when controlling for expected MA growth and obtain similar results.[21]

To further validate the instrument (recentered MA growth), we demonstrate that it is uncorrelated with expected MA growth, as shown in Panel A of Figure D1. Additionally, we compare the regression results of unadjusted MA growth and recentered MA growth on geographic variables in Panel B. While the unadjusted MA growth is significantly correlated with geographic characteristics, recentered MA growth shows no significant relationship with them. This indicates that recentering effectively removes the systematic geographic variation in market access, thereby bolstering confidence in our assignment process assumption.

Excluding confounding factors. We consider the potential influence of officials' personal characteristics in Table D2 and Table D3 and fiscal conditions in Table D4 to control for differences in local incentives or bargaining power. 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. We then exclude the influence of cultural norms and social organizations in Table D7, and natural geographical factors in Table D8. All results consistently indicate the robustness of our baseline findings.

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 D9. These tests account for various scenarios where centripetal power may be weaker, and the results remain consistent with our baseline. Combined with Table C5, 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 D2, Figure D3, Table D10, and Table D11 show that our conclusion is robust to alternative time windows, clustering methods, dependent variables, and distance measurements.

21. Notably, the estimates on $MA\ growth$ are quantitatively similar to those in column (6) using recentered-IV, which implies the exogeneity of the instrument. This is motivated by Borusyak and Hull (2023). We particularly thank Peter Hull for pointing this out and recommending us to conduct a balance test to further validate the instrument.

6 Possible Channels and Alternative Explanations

6.1 Agglomeration or Spillover Effect?

In Hypothesis 1 and Hypothesis 2, we outline two competing mechanisms to explain our main findings. In this subsection, we provide evidence that the increase in protests following the HSR in provincial capitals is primarily driven by A-to-NA migrants rather than by resource spillovers to peripheral areas. We investigate these two channels in two steps. 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 show that HSR indeed 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, and the effect is indeed stronger in areas where labor discrimination is more severe. In sum, we highlight the role of A-to-NA migrants in driving our main findings.

6.1.1 Economic and Population Agglomeration

We first examine how HSR affects population flows. Column (1) in Table 5 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. To address concerns about non-parallel trends in population flows between provincial capitals and local prefectures, Figure E1 plots the dynamic effects of HSR opening on float ratio across prefectures at varying distances from their capitals. The figure 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, rather than causing spillover effects.

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 5, 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 2, 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 E1 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 (and worsened conditions in central areas) proposed by Hypothesis 2. Instead, increasing job opportunities in central areas induced by HSR could attract more migrants, leading to ever-rising agglomeration, which validates Hypothesis 1.

TABLE 5 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 5 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.

We further test the spillover effect using firm-level data. As mentioned by Gumpert, Steimer, and Antoni (2022), firm establishments may grow in peripheral regions with HSR opening. Table E2 presents the impact of HSR on the scale of firms, measured by annual industrial output, using sample on Chinese large-scale industries. The results show that HSR opening increases industrial output in central regions while reducing it in remote areas. This again shows that HSR fosters resource agglomeration instead of spillover.

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.

6.1.2 The role of A-to-NA migrants

How does agglomeration contribute to protests? We argue that the increase in protests in provincial capital is driven by the influx of low-skilled A-to-NA migrants following the HSR opening. Due to economic disparities and institutional barriers, they face severe discrimination in provincial capitals, which leads to rising grievances and protests. 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-related protest drive our main results. Second, we use micro-survey data to show that HSR-related migration flows predominantly consist of low-skilled laborers. Third, we exploit regional variation in labor market discrimination to show that the effect of HSR on protest is significantly stronger in areas where migrant workers face more severe discrimination.

TABLE 6 The impact of HSR opening on protest involving migrant workers.

Definition	*Migration worker-related*		*Migration worker-Hukou-related*	
	(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 6 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.

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 6, 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

problem that they encounter (An et al., 2024). As shown in Panel A of Table 6, 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.

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.

Finally, apart from their pre-determined characteristics, we further explore the discrimination that they might face after coming to the provincial capital. The effect should be stronger in regions with higher levels of discrimination, where migrant workers are more likely to face unequal treatment. We use various methods to measure *objective* differences, namely, dialect, population, and economic gaps between the provincial capital and other prefectures. With greater dialect differences or larger economic disparity, migrant workers are more likely to experience greater discrimination and a lower sense of belonging in the provincial capital. Table E4 (columns (1)-(6)) presents the results, indicating that after HSR openings, provincial capitals with a higher likelihood of discrimination experience a greater increase in protests. Furthermore, we directly measure the degree of discrimination at the individual level using the CSS survey data. We aggregate the perceived discrimination among rural and urban *hukou* holders at the prefecture level and calculate the difference by subtracting the urban perception from the rural perception. A larger difference indicates that rural residents perceive greater discrimination compared to their urban counterparts, reflecting higher *subjective* feelings of discrimination. Columns (7) and (8) of Table E4 use this indicator to measure discrimination and yield a similar pattern. In sum, the agglomeration effects appear to be the primary driver of our baseline findings.

6.2 Accountability or Deterrence Effect?

Aside from its economic impact, HSR serves as an extension of political power, enhancing state presence in remote areas, as discussed in Hypothesis 3 and Hypothesis 4. In this subsection, we find little evidence that increased government accountability is driving our observed effects. Instead, we provide suggestive, albeit modest, evidence consistent with a deterrence effect.

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 accountability 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*).[23] However, columns (1) and (2) of Table 7 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 accountability effect (Hypothesis 3) fails to explain our findings.

We then provide some suggestive evidence for the deterrence effect proposed in Hypothesis 4, testing the influence of HSR opening on political trust. In CFPS, respondents are asked to rate their trust in local government officials on a scale from 0 to 10. If residents in remote regions benefit from less corruption due to HSR opening, they should develop higher political trust. Conversely, if they are deterred by state power and reduce protest to avoid punishment, their political trust may be lower. Column (1) of Table E3 shows that residents in peripheral regions experience trust loss while those in central regions trust the government more. This accords with the *deterrence effect*. As migrants play a crucial role in driving our effects, we partition the sample into locals and migrants and estimate their change in political trust respectively. Those migrants living in the capital should experience the greatest loss of political trust, as they face severe discrimination due primarily to the *hukou* system and poor living conditions as mentioned in subsection 5.3. Locals in central regions, however, may benefit from the agglomeration effect and develop positive attitudes. If the deterrence effect holds, locals in remote areas will also lose trust, while migrants may show ambiguous result. Results in columns (2)-(3) corroborate our theory: HSR opening exerts a strong negative influence on the political trust of migrants in the center, with a small positive impact on locals in the center; locals in remote areas lose trust slightly due to stronger deterrence. We use a binary indicator for high trust in columns (4)-(6) and obtain consistent findings.

Finally, we employ the ISCIU dataset to directly 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 E5 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.

23. 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).

In summary, the extension of HSR network reshapes political geography by both its economic agglomeration force and political deterrence power. On one hand, it facilitates within-province migration, which not only fosters economic agglomeration but transfers protests from remote areas to central regions. On the other, it might enable the state to extend its coercive reach into remote areas, repressing potential unrest. Migrants from remote areas to the center are prone to both effects, thus presenting the most reduction in political trust. Combined with the above evidence, we may conclude that these A-to-NA migrants contribute mainly to our findings.[24]

TABLE 7 Falsifying accountability effects

Dependent variable	(1) *Investigate*	(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 7 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.

24. Admittedly HSR opening could foster high-skilled teamwork (Dong et al., 2020), it can hardly affect our findings as A-to-NA migrants are disproportionately more than high-skilled workers like researchers. Also, they suffer particularly from improper treatment in provincial capitals thus being inclined to protest, while researchers enjoy much higher social status and economic gains.

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 land compensation. These protests may 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 E6 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. In summary, HSR opening does not trigger instant protests that may be related to HSR construction or route choices. This also rules out the concern that the opening date may represent a moment of political spotlight and citizens take advantage to manifest their grievances. Rather, they develop grievances over time (when faced with discrimination), which makes the rise of protests in central regions highly persistent, as shown in Figure 3.

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 upper-level 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*") and examine its variation with HSR opening. Figure E3 shows that these protests are hardly influenced by HSR opening. This excludes the possibility that protesters are merely commuting to central regions for petitioning.

Changes in public good provision. One possible explanation is that the provincial government may provide local governments with more transfers to support HSR construction, redistributing resources from the capital to peripheral regions. To rule out this hypothesis, we test the influence of HSR opening on the provision of public goods in Table E7, 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. The data primarily come from the *Chinese City Statistical Yearbook*. We find no evidence that HSR opening could increase the provision of these public goods.

7 Conclusion and Discussion

This study differentiates itself by focusing on the spatial heterogeneity of HSR's impact, offering a more comprehensive understanding of the underlying dynamics. We focus on HSR opening in China, highlighting its spatially heterogeneous effect: mitigating protests in remote regions while increasing them in central areas. Using monthly-level data and data on labor protests yield similar results. Moreover, results using TWFE estimator, heterogeneity-robust estimator, and recentered-IV approach produce consistent findings. We then integrate the traditional monocentric urban model with individual rational participation in protests, providing four possible explanations for this effect. Empirical evidence primarily endorses the *agglomeration effect* and excludes the *spillover effect*, emphasizing that migration from peripheral areas to central regions drives up protests. Notably, A-to-NA migrants contribute primarily to our observed effects. Meanwhile, we provide suggestive evidence for the *deterrence effect* but against the *accountability effect*. The two channels comprehensively map out HSR's effect: while increased state presence may reduce protest participation in peripheral regions, enhanced connectivity draws residents to the center and leads to rising protests, eventually reshaping the spatial distribution of protests.

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. Also, embedding our setup within a general framework with multiple centers or endogenizing the centripetal force of centers is an exciting avenue for future research.

It should be noted that the spillover effect we ruled out here is different from those established in the literature (Donaldson and Hornbeck, 2016; Donaldson, 2018), as we adopt a different definition of "spillover". In previous studies, the spillover effect is inversely related to transport cost, weakening with distance. In contrast, our spillover effect grows stronger with increasing distance. Hence, our findings are generally consistent with extant research, particularly the Fogelian view that transportation infrastructure by itself does not always promote economic growth, whose effect could vary by regions. In addition, we recognize the possible economic growth or expanded market access due to HSR openings, yet it may not present the particular geographical pattern that is crucial to bring about the change in protest geography documented in our study. In this sense, we extend the work of Banerjee et al. (2020) that focuses primarily on the effect of infrastructure for the *average location*. We show that HSR opening could bring about heterogeneous economic implications in *different locations*, which is key to engendering the changes in protest geography.

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.

Afridi, F., S. X. Li, and Y. Ren (2015). Social identity and inequality: The impact of china's hukou system. *Journal of Public Economics 123*, 17–29.

Ahlfeldt, G. M. and A. Feddersen (2018). From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography 18*(2), 355–390.

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

Allen, T. and C. Arkolakis (2022). The welfare effects of transportation infrastructure improvements. *Review of Economic Studies 89*(6), 2911–2957.

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.

Asher, S. and P. Novosad (2020). Rural roads and local economic development. *American Economic Review 110*(3), 797–823.

Banerjee, A., E. Duflo, and N. Qian (2020). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics 145*, 102442.

Barone, G., A. D'Ignazio, G. De Blasio, and P. Naticchioni (2016). Mr. rossi, mr. hu and politics. the role of immigration in shaping natives' voting behavior. *Journal of Public Economics 136*, 1–13.

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.

Beraja, M., A. Kao, D. Y. Yang, and N. Yuchtman (2023). Ai-tocracy. *Quarterly Journal of Economics 138*(3), 1349–1402.

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

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.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2020). Geography, transportation, and endogenous trade costs. *Econometrica 88*(2), 657–691.

Brinkman, J. and J. Lin (2024). Freeway revolts! the quality of life effects of highways. *Review of Economics and Statistics 106*(5), 1268–1284.

Bursztyn, L., D. Cantoni, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2021). Persistent political engagement: Social interactions and the dynamics of protest movements. *American Economic Review: Insights 3*(2), 233–250.

Calamunci, F. and J. Lonsky (2024). The road to crime: An unintended consequence of the interstate highway system. *Economic Journal*, forthcoming.

Calderon, A., V. Fouka, and M. Tabellini (2023). Racial diversity and racial policy preferences: the great migration and civil rights. *Review of Economic Studies 90*(1), 165–200.

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.

Cantoni, D., D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2019). Protests as strategic games: experimental evidence from hong kong's antiauthoritarian movement. *Quarterly Journal of Economics 134*(2), 1021–1077.

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.

Castells-Quintana, D., M. del Pilar Lopez-Uribe, and T. K. McDermott (2022). Population displacement and urban conflict: Global evidence from more than 3300 flood events. *Journal of Development Economics 158*, 102922.

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.

Chen, Y. and D. Y. Yang (2019). The impact of media censorship: 1984 or brave new world? *American Economic Review 109*(6), 2294–2332.

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.

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). 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.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A "market access" approach. *Quarterly Journal of Economics 131*(2), 799–858.

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., P. M. Morrow, and M. A. Turner (2014). Roads and trade: Evidence from the us. *Review of Economic Studies 81*(2), 681–724.

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 ....

Fogel, R. W. (1964). *Railroads and American economic growth*. Johns Hopkins Press Baltimore.

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.

Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg (2022). Migrants and firms: Evidence from china. *American Economic Review 112*(6), 1885–1914.

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

Karadja, M. and E. Prawitz (2019). Exit, voice, and political change: Evidence from swedish mass migration to the united states. *Journal of Political Economy 127*(4), 1864–1925.

Kung, J. K.-s. and C. Ma (2014). Can cultural norms reduce conflicts? confucianism and peasant rebellions in qing china. *Journal of Development Economics 111*, 132–149.

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

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, J. and G. Meng (2023). Pollution exposure and social conflicts: Evidence from china's daily data. *Journal of Environmental Economics and Management 121*, 102870.

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.

McAdam, D. (1999). *Political process and the development of black insurgency, 1930-1970*. University of Chicago Press.

McCarthy, J. D. and M. N. Zald (1977). Resource mobilization and social movements: A partial theory. *American Journal of Sociology 82*(6), 1212–1241.

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

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

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.

Persson, P. and E. Zhuravskaya (2016). The limits of career concerns in federalism: Evidence from china. *Journal of the European Economic Association 14*(2), 338–374.

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

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.

Redding, S. J. and M. A. Turner (2015). Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics 5*, 1339–1398.

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.

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

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.

Unfried, K., K. Kis-Katos, and T. Poser (2022). Water scarcity and social conflict. *Journal of Environmental Economics and Management 113*, 102633.

Voigtländer, N. and H.-J. Voth (2022). Highway to hitler. *American Economic Journal: Applied Economics*, forthcoming.

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 "State Infrastructure and Political Geography of Protest"

(For Online Publication)

Siyuan Fan Wenyi Lu

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 Accountability Effect or Deterrence Effect A5

A.5 Proof for existence and uniqueness A6

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

B.1 Institutional Background and Anecdotal Evidence A8

B.2 Details about Data Sources A10

B.3 Figures and Tables A11

Appendix C More on Empirical Results **A17**

Appendix D More on Robustness Checks **A29**

D.1 Recenter IV Approach, Long-difference Setting and Market Access Exposure A29

D.2 More Controls A32

D.3 Sub-sample Tests: Possible Outliers A39

D.4 Alternative Model Specifications A41

Appendix E More on Channels **A44**

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 Accountability 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, raising accountability, 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 (Accountability 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

B.1 Institutional Background and Anecdotal Evidence

***Hukou* System in China** In 1958, the CPC established the *hukou* (household registration) system in China to facilitate social control, under which every citizen was assigned a *hukou* type, agricultural or non-agricultural, and affiliated with a specific locality. Before 1978, the *hukou* served as a prerequisite for anyone to live and work in the designated locality. In addition, conversions from agricultural *hukou* to non-agricultural *hukou* or changing the *hukou* locality were strictly restricted, and the *hukou* of a newborn was inherited from its parents. Consequently, citizens could hardly migrate from one place to another.

The restrictions were partially relaxed since the Reform and Opening-up in 1978. Since then, people were gradually allowed to migrate from rural areas to urban areas without a local *hukou*. With huge rural-urban income disparity, Chinese cities witnessed a dramatic increase in the number of agriculture-to-non-agriculture migrants (A-to-NA migrants).[6] In addition, due to strict restriction on inter-provincial movement, intra-provincial migration dominates the general migration pattern in China (taking up about 55% of all migrants). However, till today, the household registration system still influences the mindset of Chinese people, forging a "rural bias" that migrants are poor, rural, and uneducated, which hinders the integration of migrant workers into urban life.

Despite becoming more relaxed after 2013, the *hukou* system still prevents most A-to-NA migrants from receiving equal treatment vis-à-vis their urban counterparts. For instance, they receive fairly unjust treatment in children's education, parents' healthcare, and housing for themselves, since enrollment in local schools, access to public health care, urban pension plans, and subsidized housing is all conditional on being a local urban dweller (Imbert et al., 2022; An et al., 2024). What's more, they face discrimination in the labor market (Afridi, Li, and Ren, 2015). Most of them overwork for long, with poor living conditions, a few of whom suffer from wage arrears that incur severe economic vulnerability, as shown in the 2016 migrant worker monitoring survey report.[7] Faced with unequal treatments, they may develop strong grievance and protest for their own rights.[8]

6. The proportion of A-to-NA migrants to all migrants is at least 76% in 2000-2020. See Table 1 of Tombe and Zhu (2019) or China's seventh national census for details.

7. The report shows the demographic features of A-to-NA migrants: 65.5% of them are male; 73.6% have less than 9 years of education. It also shows their labor protection condition: 64.9% of them have never signed a labor contract; 84.4% work over 44 hours per week, with 8.7 working hours per day; 0.84% (2.37 million) have unpaid wages, with average wage arrear being 11,941 RMB, equal to their three months' earnings (Considering daily expenditure, such wage arrear may lead to severe economic vulnerability of their families.). As for living conditions, 62.4% of them rent a room, with the average housing area being $19.4m^2$; nearly half of them have no washing machines or fridges; 13.5% even have no water access. See https://www.stats.gov.cn/sj/zxfb/202302/t20230203_1899495.html for details.

8. Protests of A-to-NA migrants due to wage arrears are prevalent in China. Notable examples include the Ja-

Protest of A-to-NA Migrants HSR opening has expedited the migration of rural populations to urban centers. However, as mentioned before, the persistent urban-rural divide and poor labor protection systems have engendered many unjust treatments for A-to-NA migrants. The deeply ingrained discrimination of the *hukou* system, coupled with inadequate institutional support, may have provoked these migrants, who are drawn to the cities by convenient transportation, to engage in protests in central regions.

A-to-NA migrants in China frequently confront systemic discrimination that permeates various aspects of urban life. The *hukou* system undermines the socioeconomic status of migrant workers by capping their income, restricting their access to social insurances, and eventually hindering their integration into urban life. Such institutional bias not only impacts their financial prospects but also constructs a broader social narrative that positions them at severe disadvantage, nurturing an environment conducive to dissatisfaction and collective resistance.

Further, the plight of migrant workers is intensified by the social stigma and exclusion they encounter in urban environments. Despite easier access to cities fostered by HSR, poor labor protection and inequitable wages remains as HSR could by no way change their *hukou* status, which leads to high levels of dissatisfaction. Their initial hope-filled migration is gradually overshadowed by the growing awareness of daunting urban-rural disparities. For instance, *Xiaoyu*, a male migrant worker from Sichuan province, despite changing jobs multiple times and moving from one city to another, found no substantial improvement in his life. This reflects limited upward mobility in Chinese society and the marginalized status of A-to-NA migrants (Rho, 2023). Therefore, while moving to central regions with hope, most of the migrants would soon recognize their poor status and face numerous unjust treatments. Such ongoing mistreatment, alongside the recognition of the systemic roots of their struggles, can ignite strong sense of relative deprivation, leading to conflicts and protests to voice their grievances.

Eventually, the uprisings of A-to-NA migrants may transcend the realm of economic affairs, becoming a resounding call for social acknowledgment and equal rights. Rho (2023) notes that "Social grievances take center stage over economic ones in turning the tide of labor discontent among the rank and file". The ferment of their discontent is not just a reaction to personal tribulations but a profound awakening to the systemic rural-urban disparity of the Chinese society, which propels them toward unified action for equal treatments. Their protests, while initially oriented to personal interests, may become a testament to the human spirit's unwavering quest for equality and respect, even in the shadow of inequality and injustice under authoritarian governance.

sic incident in 2018. See https://www.theguardian.com/world/2018/nov/12/ten-student-activists-detained-in-china-for-supporting-workers-rights for detail.

B.2 Details about Data Sources

CMDS. In the appendix, we also employ the China Migrants Dynamic Survey (CMDS) dataset for 2011–2016 to examine whether the population agglomeration in provincial capitals following HSR openings is primarily driven by the inflow of rural-to-urban migrants. The CMDS has been conducted annually by the China National Health Commission since 2009, targeting migrants working in over 300 cities across 31 provinces in China. We leverage the CMDS to examine the demographic characteristics of migrants and to shed light on the underlying mechanisms. With a large sample size (approximately 1 million migrant records from 2011 to 2016), the survey is representative of Chinese migrants and has been widely used in academic research (An et al., 2024). We use the full sample of CMDS data from 2011 to 2016, which includes detailed information on each respondent's background, average years of education, proportion of the agricultural population, male proportion, average age, and proportion of Han ethnicity among the floating population in provincial capitals over the years.

CFPS. The China Family Panel Studies is a nationally representative survey covering 25 Chinese provinces, conducted biannually by the Institute of Social Science Survey at Peking University. We leverage two sets of variables from CFPS. First, CFPS collects data on respondents' satisfaction with local government performance (rated on a 5-point scale) and their trust in government officials (rated on a 10-point scale). Second, CFPS documents respondents' emotions (subjective attitudes), including whether the respondents feel down, have difficulty in functioning, the degree of loneliness, the level of depression, and the degree of concern about the severity of current corruption issues. These variables facilitate our analysis on the influence of HSR opening of local residents.

CSS. The Chinese Social Survey (CSS) is a nationwide, large-scale survey launched by the Chinese Academy of Social Sciences. It includes a wide range of questions about subjective attitudes on social/political issues. The survey asks respondents to what extent they consider the discrimination against rural residents is severe in China, which helps us measure discrimination against migrants.

ISCIU. The Ideology Survey of Chinese Internet Users (ISCIU) is commonly used to study the ideological leanings of Chinese people, conducted Ma and Wang (2015). Notably, ISCIU includes detailed questions on political expression, which is rare in Chinese surveys.

B.3 Figures and Tables

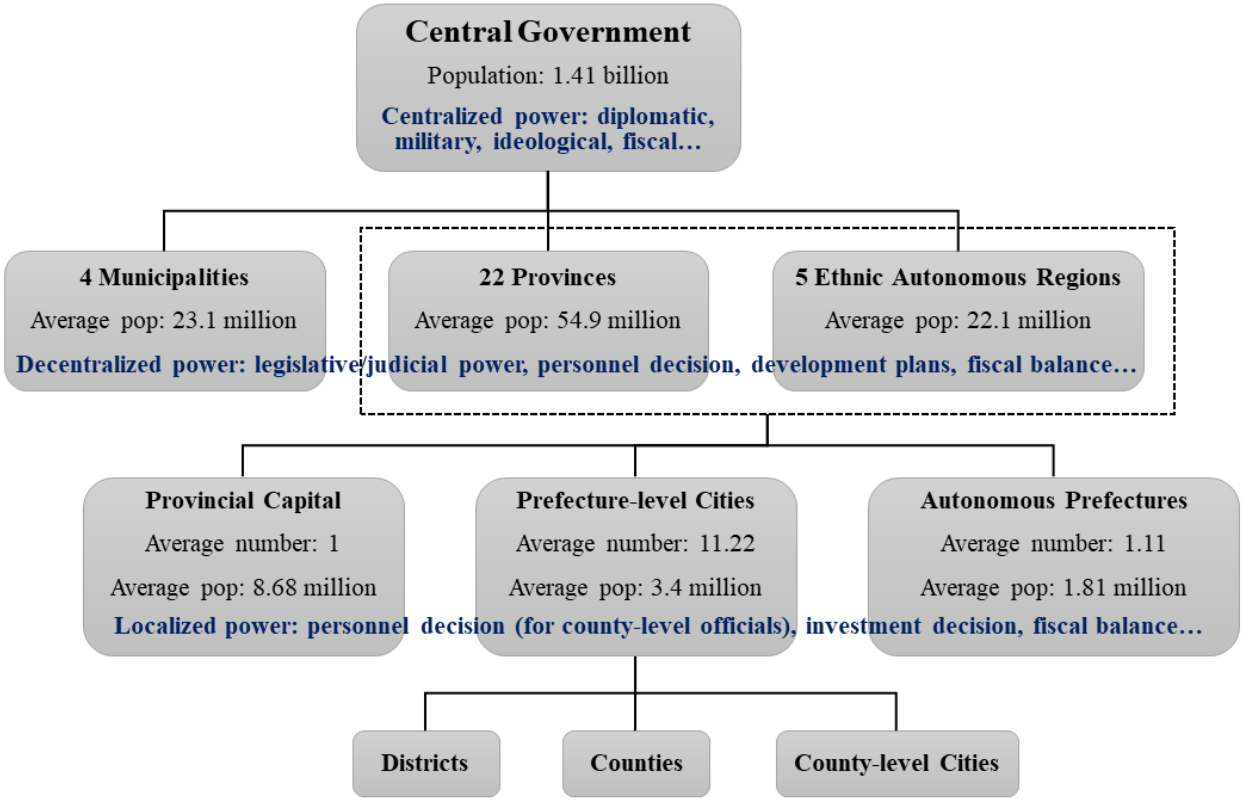


FIGURE B1 The Divisions of Chinese Bureaucracy.

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).

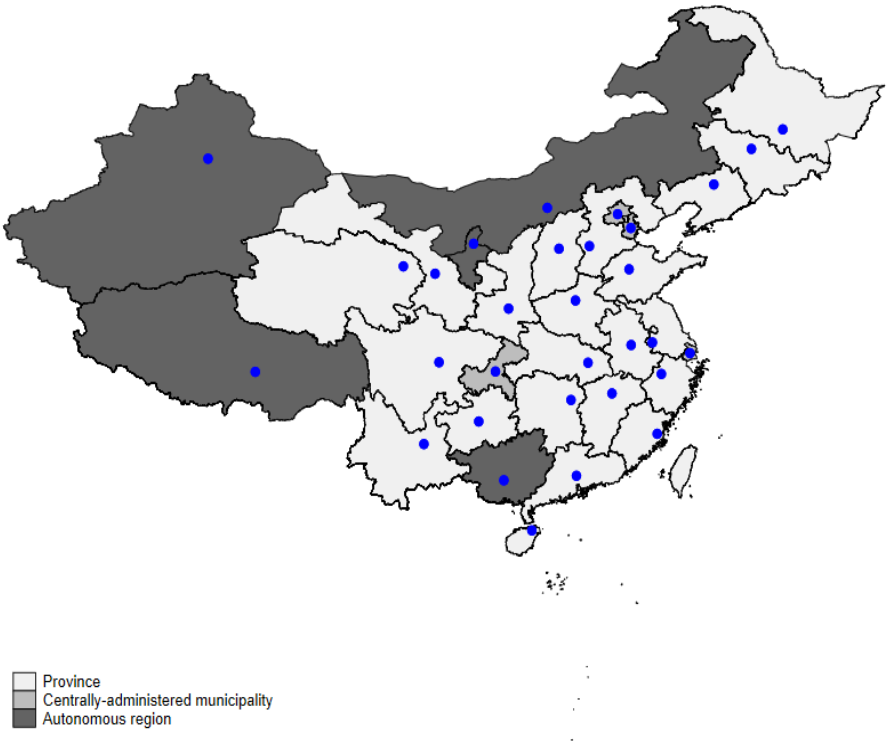


FIGURE B2 The Spatial Distribution of Provincial Regions.

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.

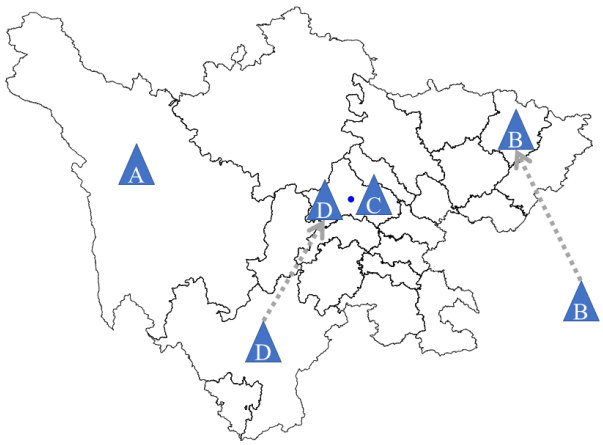


FIGURE B3 A sketch of different groups of people.

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.

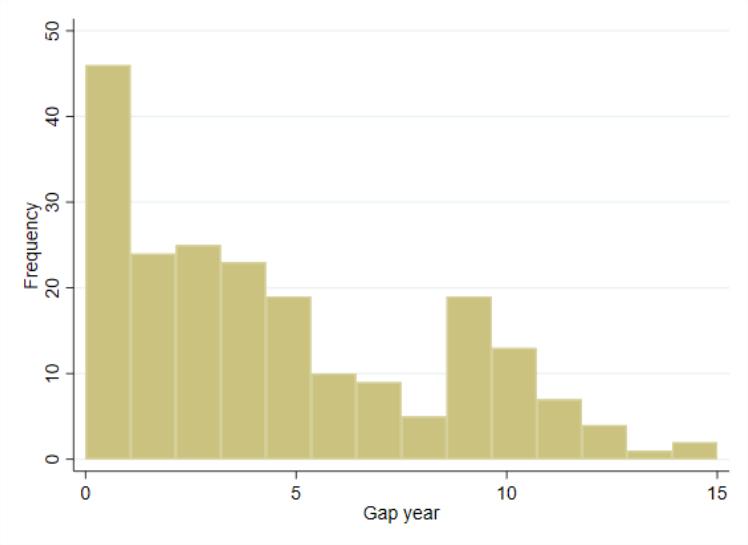


FIGURE B4 The HSR opening time gap between outer regions and capitals.

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.

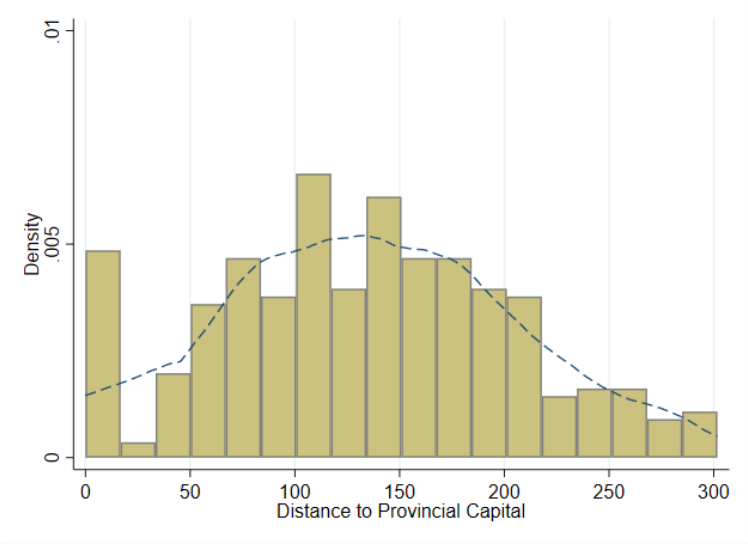
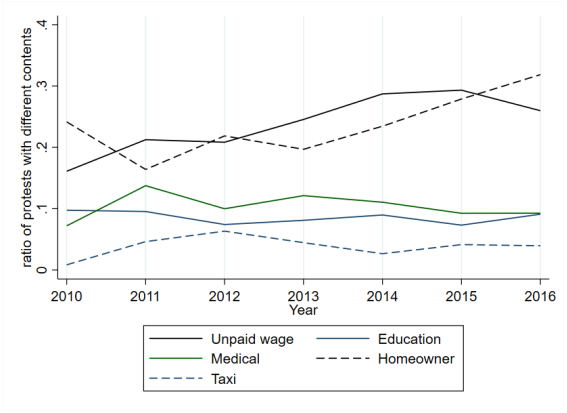
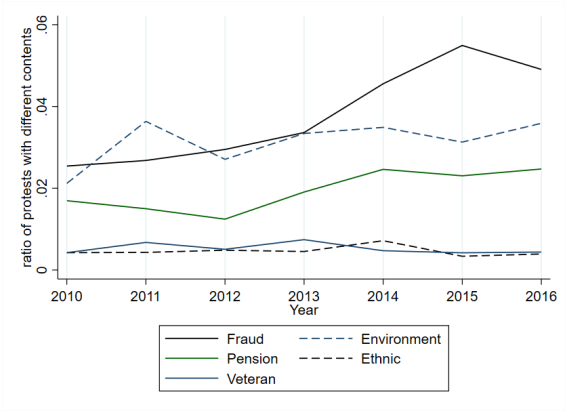


FIGURE B5 The density of geographical distances to provincial capital.

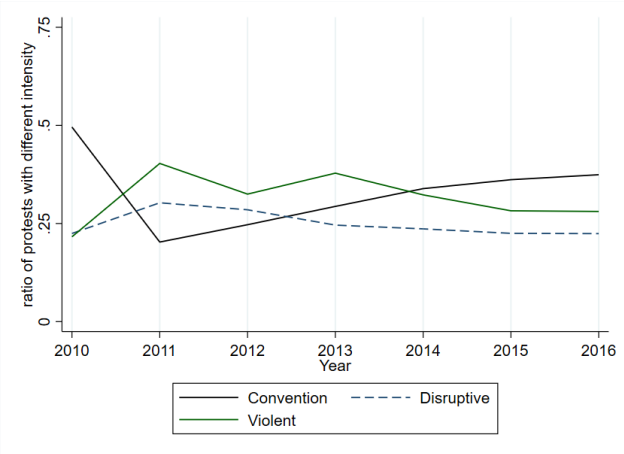
Notes: This figure presents the distribution of the geographical distance between prefectures and the provincial capital in China. The data for municipalities is excluded in our sample.



(A) **The proportion of protests of different contents (1).**

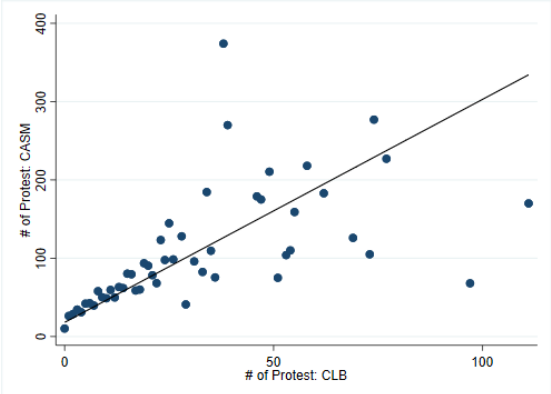


(B) **The proportion of protests of different contents (2).**

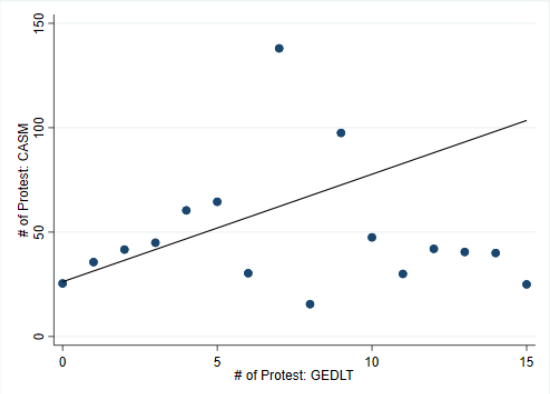


(C) **The proportion of protests of different intensity.**

FIGURE B6 The proportion of protests with different contents and intensity.



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



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

FIGURE B7 The Correlation between the Number of Protests in Different Datasets.

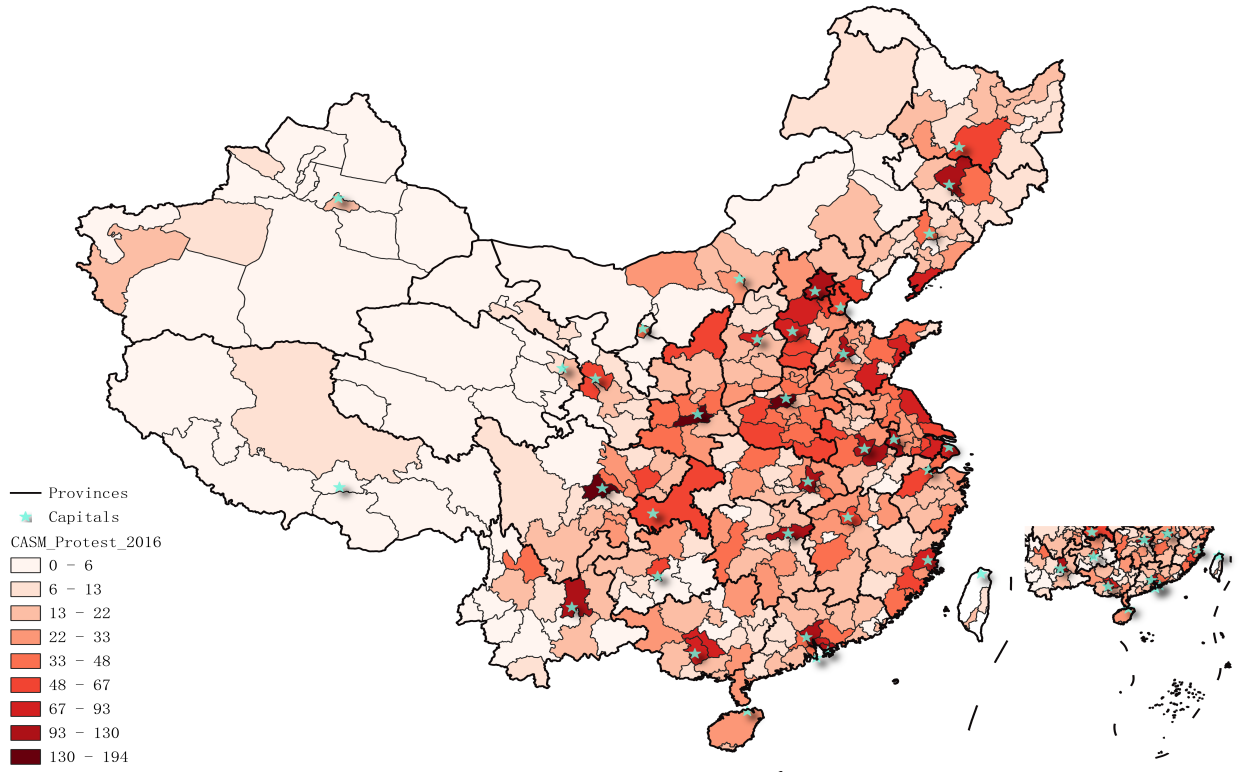


FIGURE B8 The spatial distribution of protest events in 2016.

Notes: This figure presents the spatial distribution of protest events in 2016 (the end of our sample period). The color signal the frequency of protest events in each prefecture, with deeper red color representing more protests while lighter color for fewer protests. The blue stars are the capital cities of different provinces.

TABLE B1 Relationship between HSR and Internet Penetration

Dependent variable	*Internet Penetration*			
	(1)	(2)	(3)	(4)
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)
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Year FE	✓		✓	
Province-year FE		✓		✓
Observations	1,902	1,895	1,889	1,882
R-squared	0.782	0.829	0.783	0.829

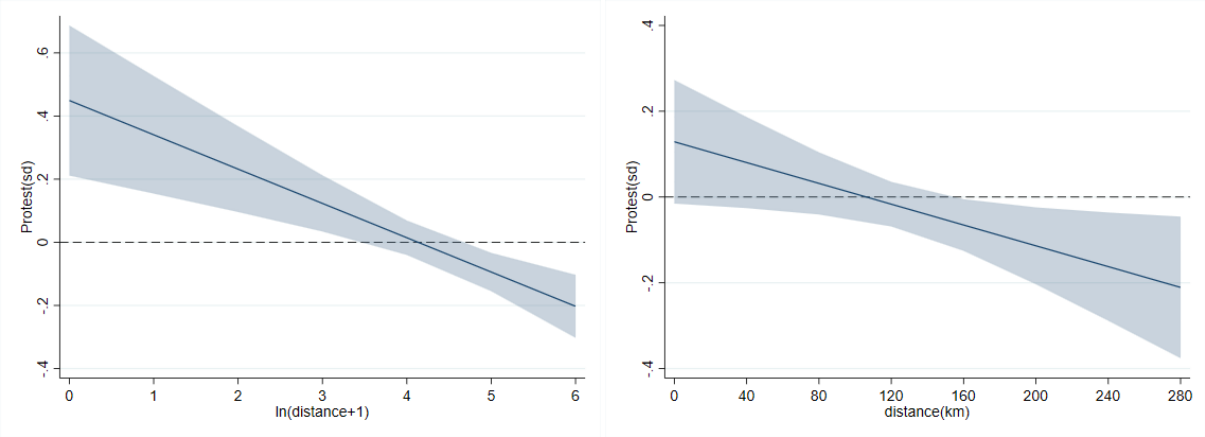
Notes: Table B1 reports the relationship between HSR access and internet penetration. The dependent variable is the standardized number of households with internet access in each prefecture (*Internet Penetration*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 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
Panel C: Prefecture-month Outcomes					
Strike_month	31,008	0.313	0.875	0	15
Strike_month	31,008	0	1	-0.358	16.776

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

Appendix C More on Empirical Results

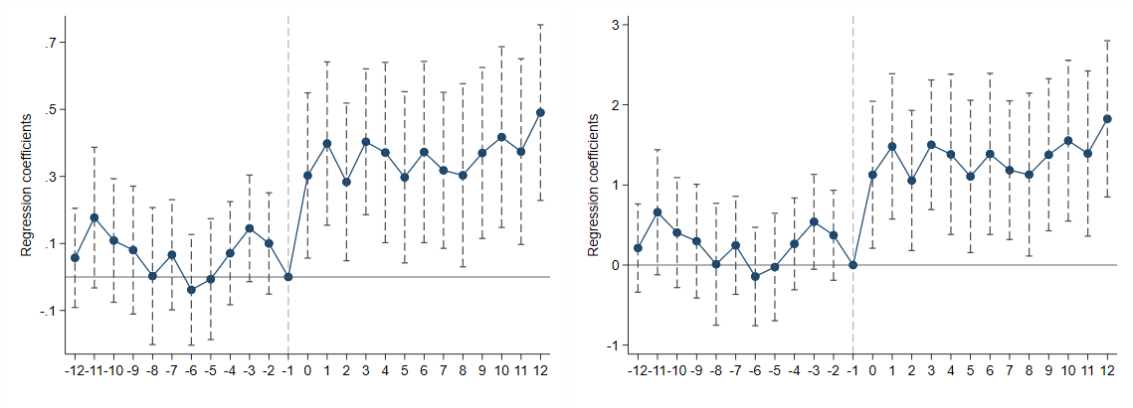


(A) **The heterogeneous impact for different distance (logged) from provincial capital.**

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

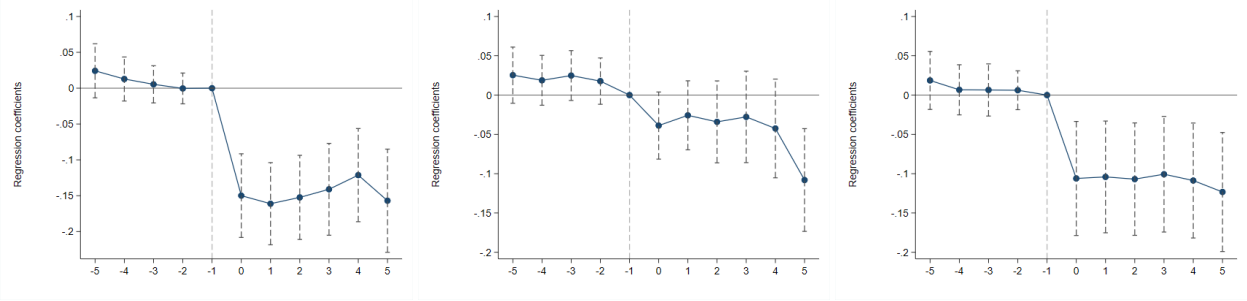
(B) **Pre-trend tests for *HSR* term using the number of protests.**

(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 C2 Pre-trend tests using prefecture-month level data.

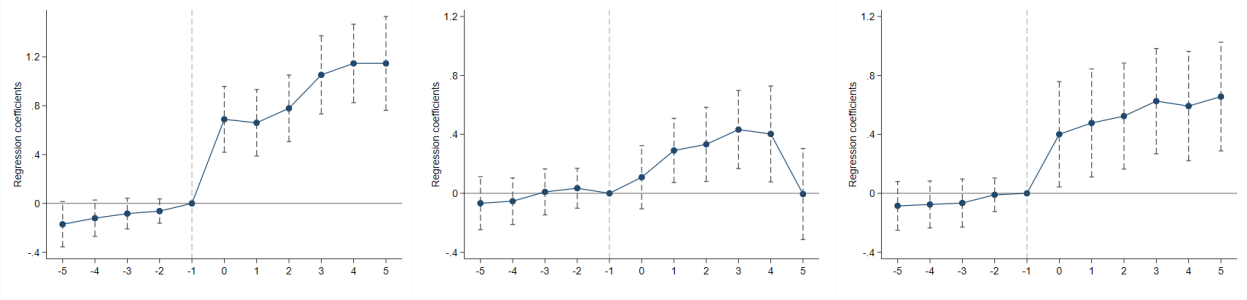
Notes: This figure plots pre-trend tests using prefecture-month level data. The dependent variable is the standardized number of protests at the prefecture-month level (*Protestsd_month*) in Panel A and Panel C, while we use the actual number of protests (*Protest_month*) in Panel B and Panel D. Panel A and Panel B test for single *HSR* term. Panel C and Panel D test for the interaction term *HSR*×*lndistance* in equation (2); Each point represents a month, with the leftmost and rightmost points representing all dates before the 12th month and after the 12th month, respectively. All panels control for Province-year-month, predetermined-year, and prefecture-month 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 (Interaction term)

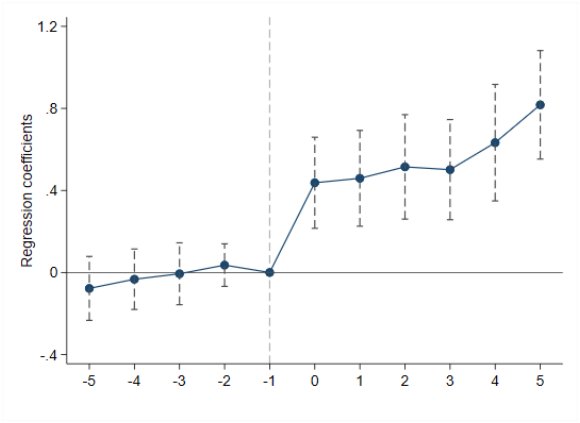
Notes: This figure plots pre-trend tests for interaction term $HSR \times lndistance$ in equation (2) using different classifications based on intensity. The dependent variables for Panels A-C are standardized prefecture-level protest counts of different intensities, namely conventional, disruptive, 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.



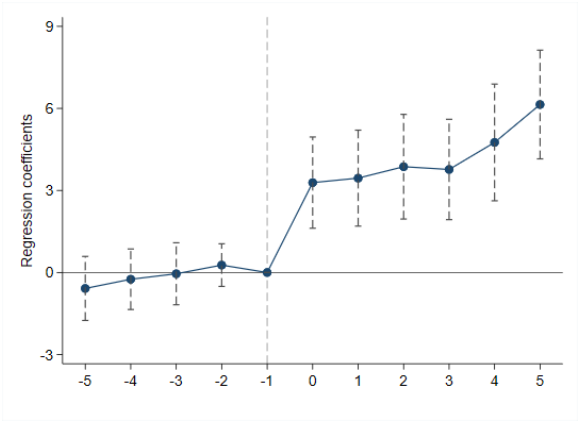
(A) **Convention events.** (B) **Disruptive events.** (C) **Violent events.**

FIGURE C4 Pre-trend tests for different protest issues (HSR term)

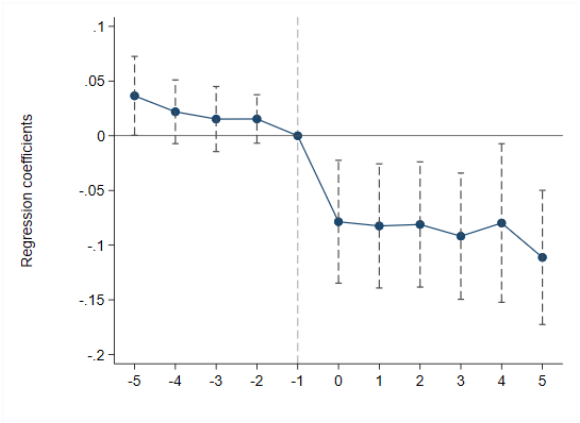
Notes: This figure plots pre-trend tests for single term *HSR* in equation (2) using different classifications based on intensity. The dependent variables for Panels A-C are standardized prefecture-level protest counts of different intensities, namely conventional, disruptive, and violent protests. All 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.



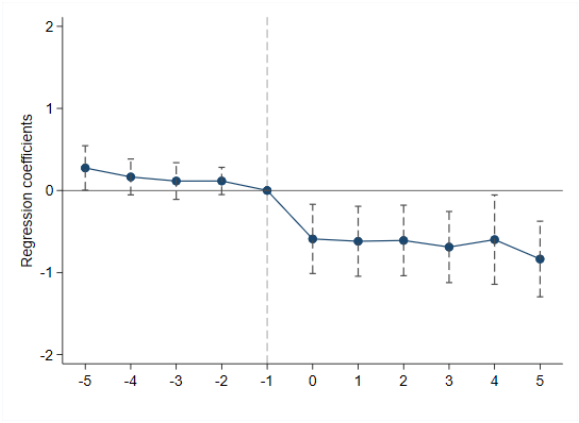
(A) **Pre-trend tests for *HSR* term using the standardized number of protests.**



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



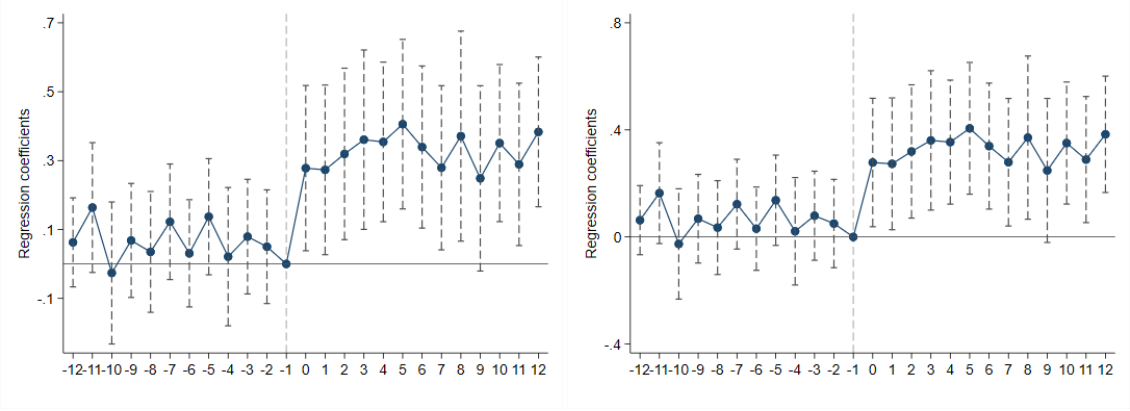
(C) **Pre-trend tests for interaction term using the standardized number of protests.**



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

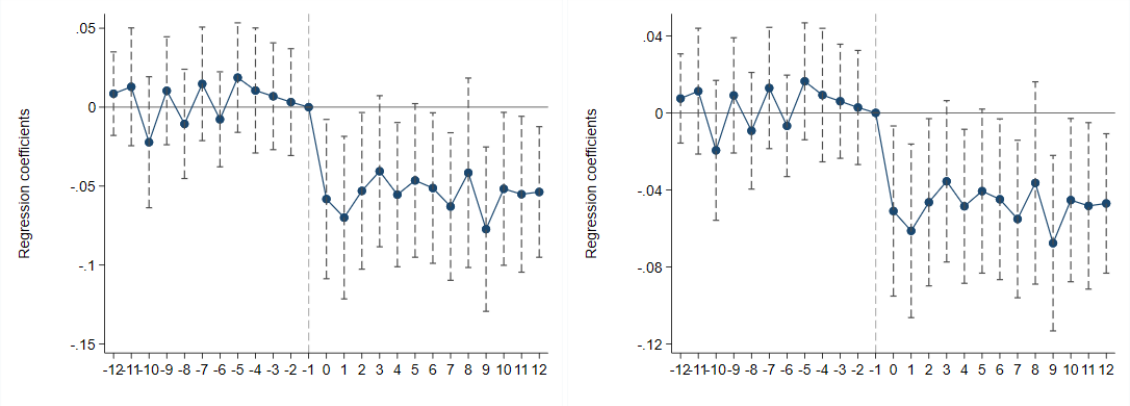
FIGURE C5 Pre-trend tests using CLB dataset.

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

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

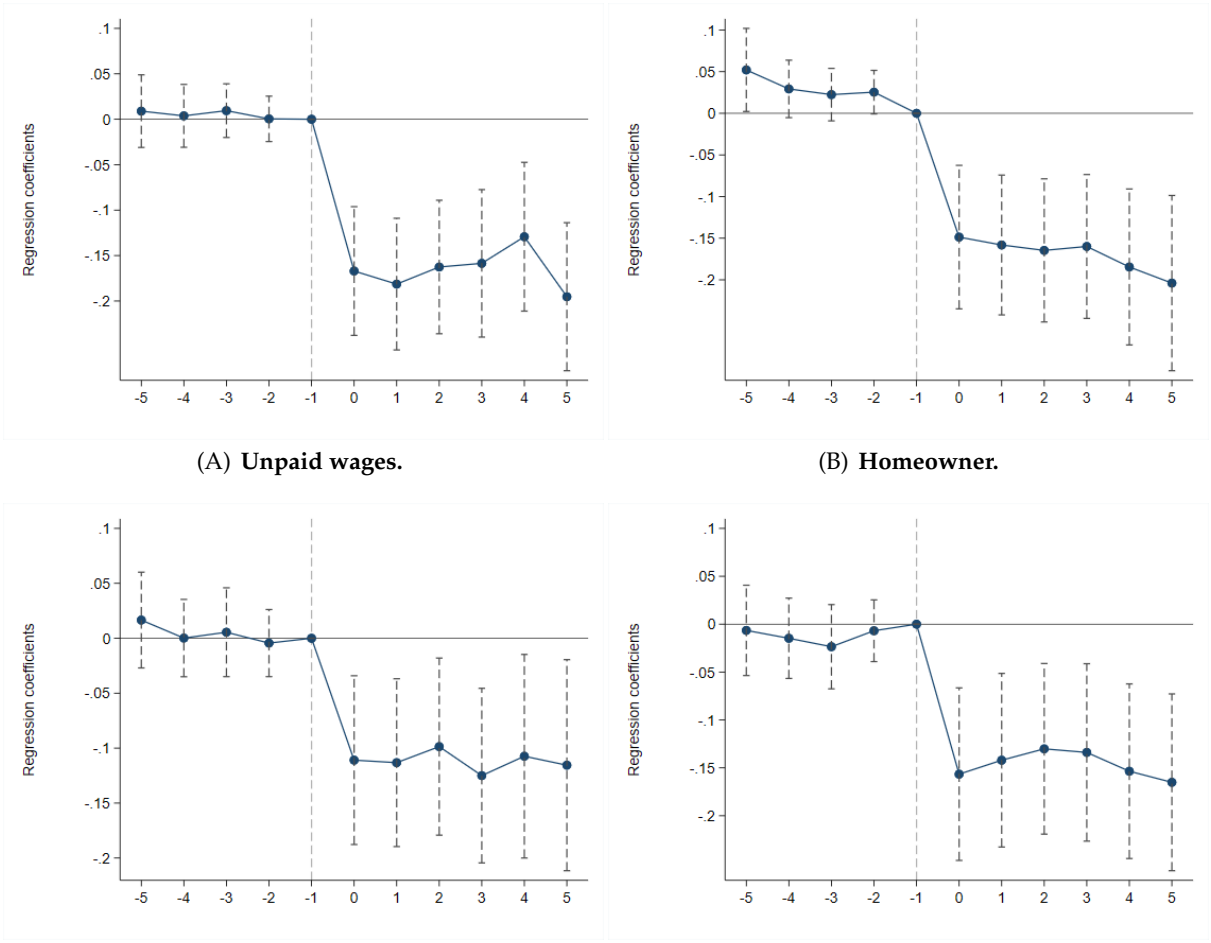


(C) **Pre-trend tests for interaction term using the standardized number of protests.**

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

FIGURE C6 Pre-trend tests using prefecture-month level data in CLB dataset.

Notes: This figure plots pre-trend tests using prefecture-month level data in CLB dataset. The dependent variable is the standardized number of protests at the prefecture-month level (*Protestsd_month*) in Panel A and Panel C, while we use the actual number of protests (*Protest_month*) 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); Each point represents a month, with the leftmost and rightmost points representing all dates before the 12th month and after the 12th month, respectively. All panels control for Province-year-month, predetermined-year, and prefecture-month fixed effects. Robust standard errors are clustered at the prefecture levels. The dotted line shows a 95% confidence interval.

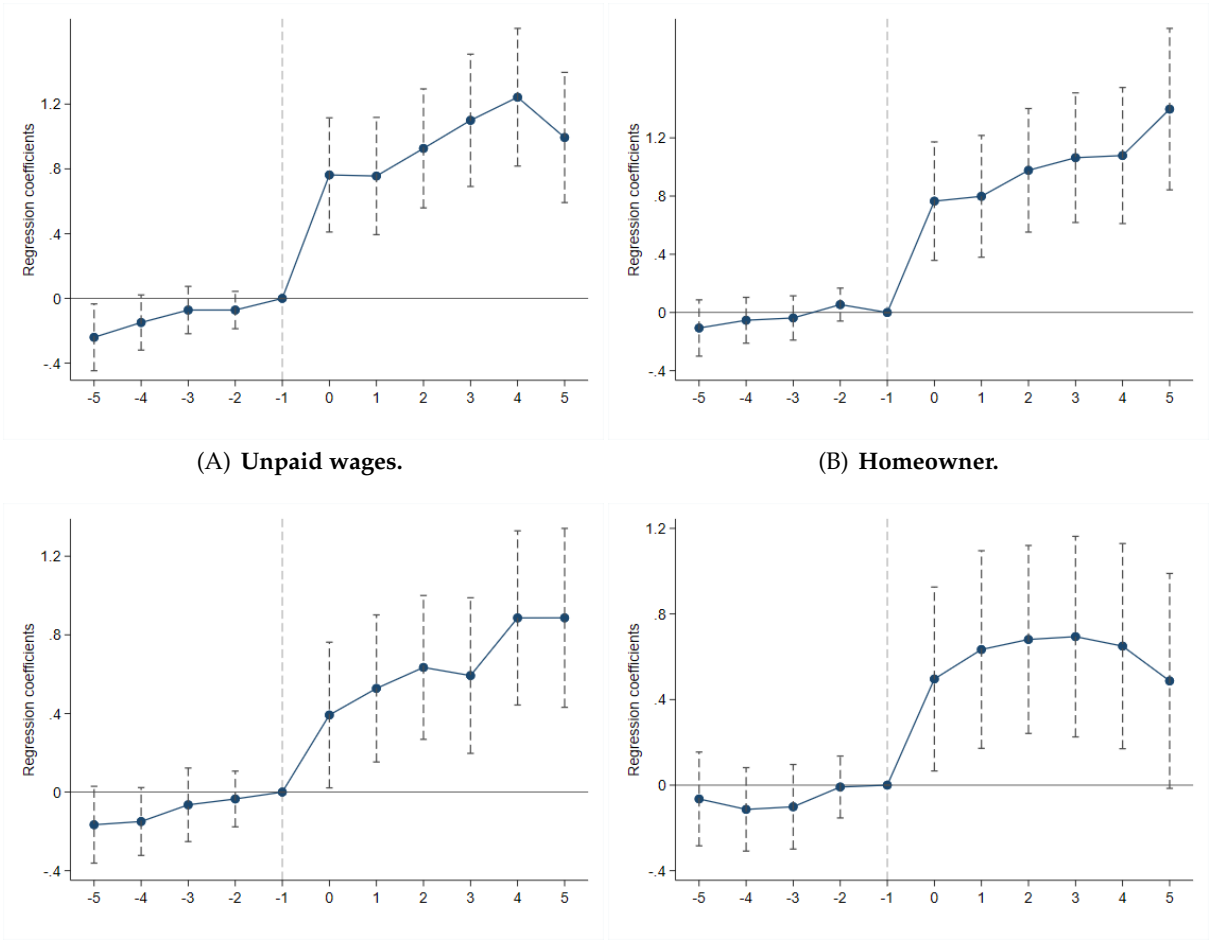


(A) **Unpaid wages.** (B) **Homeowner.**

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

FIGURE C7 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.

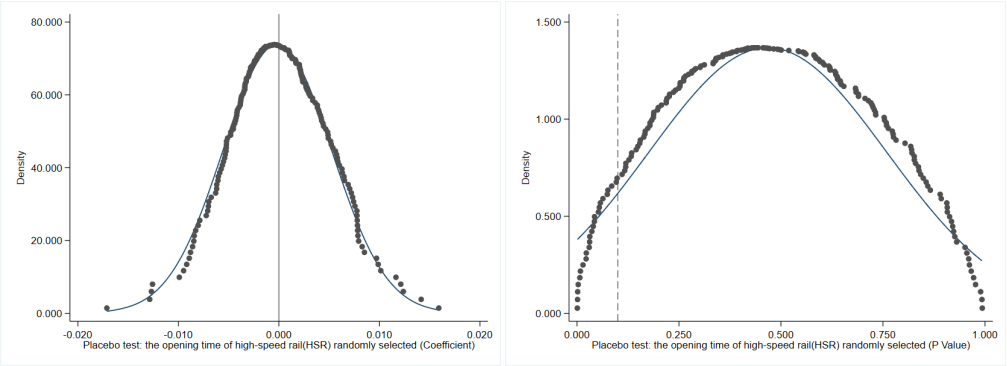


(A) **Unpaid wages.** (B) **Homeowner.**

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

FIGURE C8 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.



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

FIGURE C9 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.

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 Baseline Results using Prefecture-month Data

	(1)	(2)	(3)	(4)	(5)
HSR	0.000	0.252**	0.296***	0.317***	0.325***
	(0.039)	(0.117)	(0.112)	(0.121)	(0.124)
	[0.035]	[0.120]	[0.121]	[0.124]	[0.127]
HSR×*lndistance*		-0.063***	-0.067***	-0.072***	-0.074***
		(0.024)	(0.023)	(0.024)	(0.025)
		[0.026]	[0.025]	[0.026]	[0.027]
Predetermined-year Control	✓	✓	✓	✓	✓
Prefecture FE		✓	✓	✓	
Month FE		✓	✓		
Year FE		✓			
Province-year FE			✓		
Province-year-month FE	✓			✓	✓
Prefecture-month FE	✓				✓
Observations	22,920	22,836	22,836	22,752	22,752
R-squared	0.734	0.626	0.652	0.703	0.736

Notes: Table C3 reports the baseline regression results using prefecture-month level data. The dependent variable is the standardized number of protests at the prefecture-month level (*Protestsd_month*). *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 C4 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 C4 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.

TABLE C5 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 C5 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 C6 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 C6 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.

Appendix D More on Robustness Checks

D.1 Recenter IV Approach, Long-difference Setting and Market Access Exposure

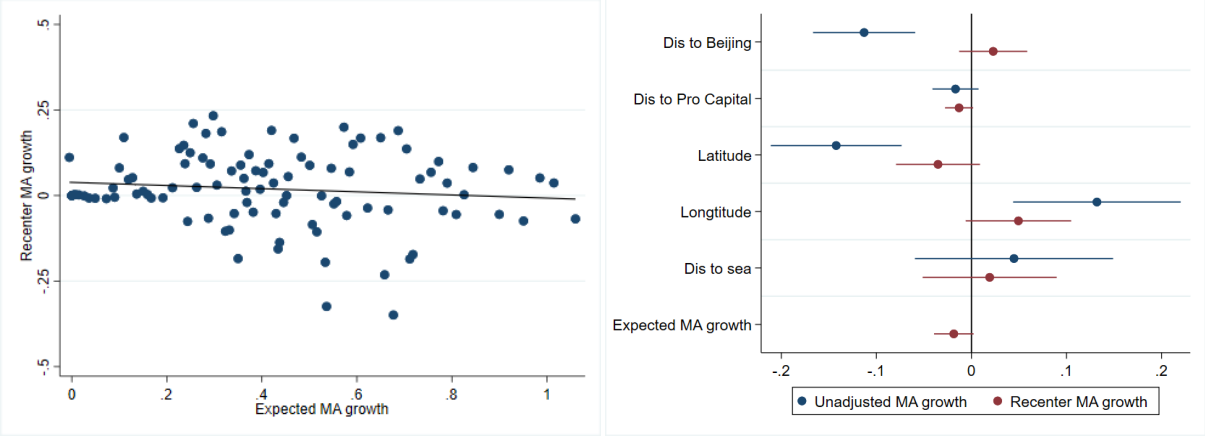
In Table D1, we construct a cross-sectional dataset to examine the relationship between changes in HSR access and changes in protest events during 2010-2017. We restrict the sample to regions without HSR access in 2010 and obtain 226 prefectures (53 of 279 prefectures had HSR in 2010). We then define 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 D1. Additionally, we quantify the systematic nature of spatial variation in MA growth.

In the first regression specification depicted in Figure D1 (indicated by blue dots), the blue regression coefficients illustrate the relationship between unadjusted MA growth and economic geography 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 D1 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 D1 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 D1 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.2 More Controls

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

Firstly, we exclude the potential influence of officials' personal characteristics. The officials who take office in their hometown ("locals") might be more adept with local governance, thus inducing fewer protests. Meanwhile, locals may spend more budget on welfare rather than infrastructure (Persson and Zhuravskaya, 2016). Our results may be driven once the above two arguments hold simultaneously. We examine such effect in Table D2. In columns (1)-(3), we add control variables indicating whether the party secretary or governor of the prefecture or province is local, showing that they pose little threat to our baseline results. In columns (4)-(5), we interact them with HSR opening and distance. The coefficients are small and ambiguous, showing possibly little effect on our results. Further, we analyze 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 D3. 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 account for the potential impact of local fiscal factors. Due to HSR construction, local governments may receive additional funding from the central government. One concern is that the change in fiscal conditions from HSR construction could affect regional infrastructure and welfare expenditures, thereby influencing protests. In Table D4, we consider the effect of local fiscal deficits and fiscal structures, finding that they pose little threat to our conclusion.

Thirdly, 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.

Fourthly, 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 E7, and find weak relation.

What's more, we address concerns on the influence of local culture and organizations. Traditional cultural norms, like Confucian culture or clan culture, may influence residents' protest

willingness by shaping their values for long (Kung and Ma, 2014). As noted by abundant research (McCarthy and Zald, 1977; McAdam, 1999), organizations may either foster or palliate protests. Meanwhile, they may influence residents' reluctance to HSR opening. In Table D7, we control for Confucian culture, the number of official organizations, clan density interacted with year dummies. The results are similar to the baseline. We also consider the possible influence of political connections. With more regional members in the Central Committee, the region may possess stronger political importance. The result in column (5) shows minor difference. We control all above in column (6), finding that these factors do not affect our conclusion.

Lastly, we exclude the influence of natural geographical factors. On the one hand, they are important factors influencing the decision-making and the construction process of HSR. On the other hand, previous studies have highlighted that adverse natural conditions may exacerbate protests (Unfried et al., 2022; Li and Meng, 2023). We control for terrain, average slope and average elevation interacted with year dummies, precipitation, PM2.5, and all above in Table D8. We find that our conclusion remains highly robust even considering these factors.

TABLE D2 Controlling for locals.

	(1)	(2)	(3)	(4)	(5)
HSR	0.467***	0.433***	0.431***	0.437***	0.473***
	(0.134)	(0.136)	(0.134)	(0.138)	(0.132)
HSR×*lndistance*	-0.109***	-0.106***	-0.107***	-0.107***	-0.111***
	(0.027)	(0.027)	(0.027)	(0.028)	(0.027)
HSR×*lndistance*×				0.041	
Province Local				(0.068)	
HSR×*lndistance*×					0.158
Prefecture Local					(0.097)
Province Local Controls		✓	✓	✓	
Prefecture Local Controls			✓		✓
Predetermined-year Control	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓
Observations	1,896	1,866	1,814	1,866	1,840
R-squared	0.900	0.905	0.905	0.905	0.901

Notes: Table D2 reports results controlling for the influence of locals. ***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. *Province Local* is a dummy variable indicating whether the birthplace of the provincial party secretary or governor (*Shengwei Shuji/ Shengzhang*) in the current year belongs to this prefecture-level city, while *Prefecture Local* indicates whether the birthplace of the municipal party secretary or mayor (*Shigwei Shuji/ Shizhang*) in the current year belongs to this prefecture-level city.

TABLE D3 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 D3 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 D4 Controlling for fiscal conditions.

	(1)	(2)	(3)	(4)	(5)
		Controlling for			
	Baseline	Deficit ratio	Lagged deficit ratio	Fiscal structure	All
HSR	0.467***	0.470***	0.469***	0.469***	0.467***
	(0.134)	(0.133)	(0.133)	(0.132)	(0.133)
HSR×*lndistance*	-0.109***	-0.109***	-0.109***	-0.109***	-0.109***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Predetermined-year Control	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓
Observations	1,896	1,895	1,894	1,895	1,894
R-squared	0.900	0.900	0.900	0.900	0.900

Notes: Table D4 reports results controlling for fiscal conditions. ***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 current fiscal deficit ratio, the lagged one-year fiscal deficit ratio, and the proportion of fiscal revenue and expenditure. Column (5) controls for all the variables mentioned above.

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.

	(1)	(2)	(3)	(4)
		Controlling for		
	Baseline	Internet penetration rate	Broadband policy	All
HSR	0.467***	0.439***	0.460***	0.430***
	(0.134)	(0.143)	(0.133)	(0.143)
HSR×*lndistance*	-0.109***	-0.103***	-0.108***	-0.102***
	(0.027)	(0.029)	(0.027)	(0.029)
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,896	1,882	1,896	1,882
R-squared	0.900	0.902	0.900	0.902

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 (3) respectively control for the Internet penetration rate and Chinese Broadband policy. Column (4) controls for both of the variables mentioned above.

TABLE D7 Controlling for local culture and organizations.

	(1)	(2)	(3)	(4)	(5)	(6)
		Controlling for				
	Baseline	Confucian culture	Organizations	Clan culture	Central committees	All
HSR	0.467***	0.623***	0.620***	0.642***	0.640***	0.629***
	(0.134)	(0.090)	(0.097)	(0.086)	(0.088)	(0.097)
HSR×*lndistance*	-0.109***	-0.141***	-0.142***	-0.146***	-0.145***	-0.143***
	(0.027)	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)
Predetermined-year Control	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓
Observations	1,896	1,616	1,616	1,616	1,616	1,616
R-squared	0.900	0.904	0.907	0.904	0.903	0.910

Notes: Table D7 reports results controlling for local culture and organizations. ***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) respectively control for the interaction terms between year dummy variables and the logarithm of the number of academies, the logarithm of the number of social organizations in 2008, the logarithm of the proportion of genealogy quantity to the total population, and the number of members in central committees (1978-2017). Column (6) controls for all the variables mentioned above.

TABLE D8 Controlling for natural geographical conditions.

	(1)	(2)	(3)	(4)	(5)	(6)
		Controlling for				
	Baseline	Average slope	Average elevation	Precipitation	PM2.5	All
HSR	0.467***	0.468***	0.471***	0.640***	0.639***	0.477***
	(0.134)	(0.136)	(0.137)	(0.086)	(0.088)	(0.139)
HSR×*lndistance*	-0.109***	-0.106***	-0.108***	-0.144***	-0.145***	-0.107***
	(0.027)	(0.028)	(0.028)	(0.018)	(0.018)	(0.028)
Predetermined-year Control	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓
Observations	1,896	1,896	1,896	1,616	1,616	1,896
R-squared	0.900	0.902	0.902	0.904	0.903	0.903

Notes: Table D8 reports results controlling for natural geographical conditions. ***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)-(3) respectively control for the interaction terms of year dummy variables with average slope and average elevation. Columns (4)-(5) respectively control for precipitation and annual average PM2.5 concentration. Column (6) controls for all the variables mentioned above.

D.3 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 D9 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.[9] 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[10], 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.

9. 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.

10. 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 D9 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 D9 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.

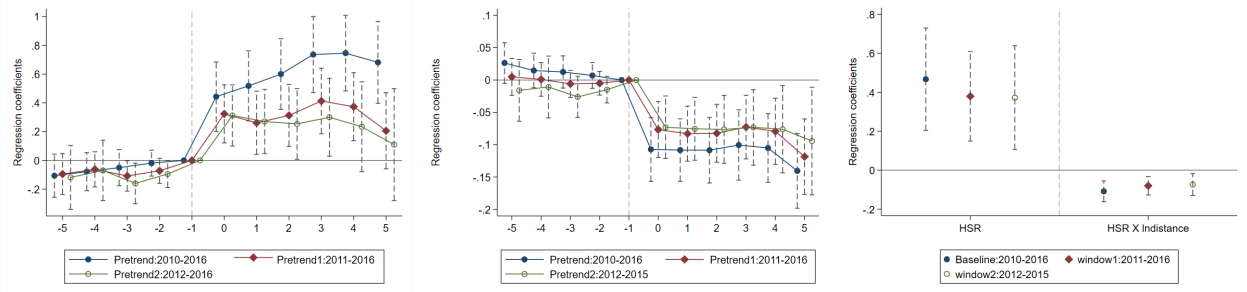
D.4 Alternative Model Specifications

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 D2 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 D3 and find similar results, for both the baseline and the parallel trend results.

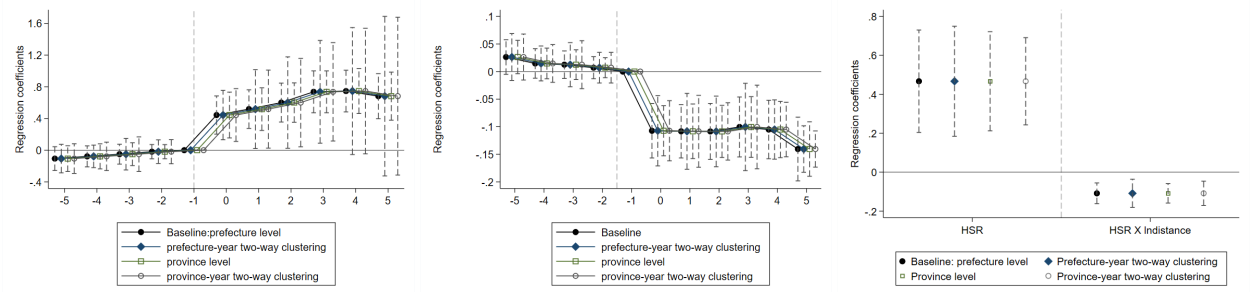
Thirdly, we examine whether different methods in the construction of dependent variables may affect our conclusion. In columns (2)-(5) of Table D10, we use the number of protests per year in each prefecture, the number of protests per million population, and the logarithm of the number of protests (plus 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 D11 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 D2 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 D3 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 D10 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

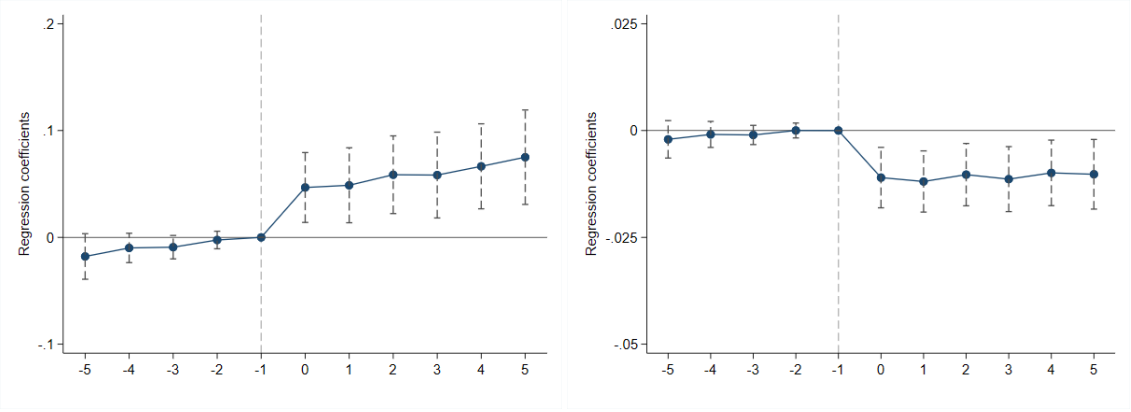
Notes: Table D10 reports results testing for alternative model specifications. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Columns (1)-(5) employ the OLS fixed effects model, while column (6) utilizes the panel Poisson model. Column (1) uses the standardized 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 D11 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 D11 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.

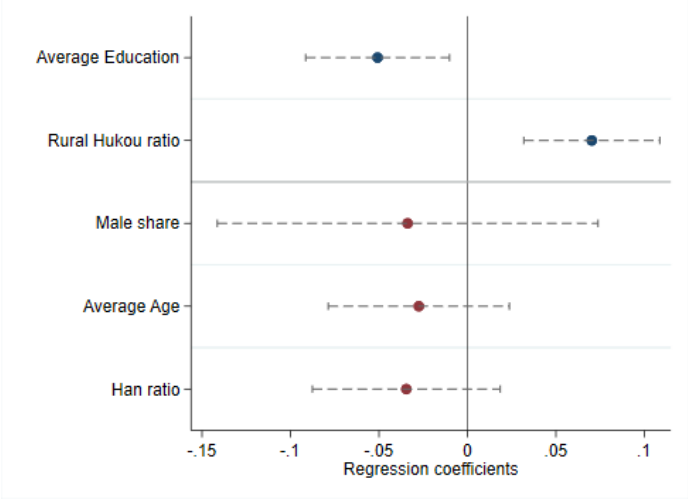
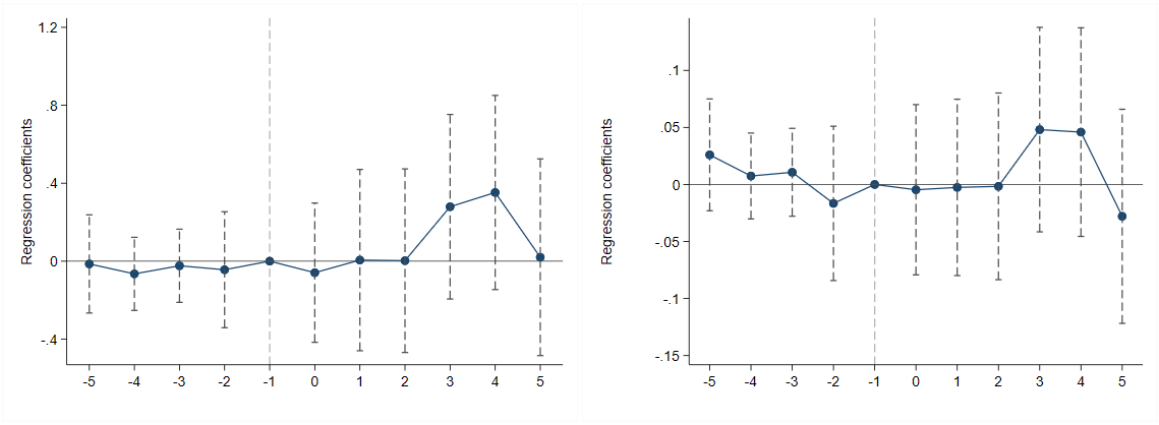


FIGURE E2 Relationship Between HSR Opening and the Average Characteristics of the Floating Population in Provincial Capitals.

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 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 E1 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 E2 The heterogeneous impact of HSR openings on the scale of enterprises.

Enterprise category	(1) *All Above-scale industries*	(2) *Domestic-funded*	(3) *Hong Kong, Macao, and Taiwan-funded*	(4) *Foreign-funded*
HSR	51.780***	45.819**	9.672**	-4.174
	(17.403)	(21.544)	(4.045)	(7.239)
HSR×*lndistance*	-13.119***	-11.447**	-2.670***	1.061
	(3.883)	(4.725)	(0.893)	(1.583)
Predetermined-year Control	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓
Observations	1,943	1,942	1,826	1,842
R-squared	0.975	0.956	0.972	0.988
Outcome mean	308.109	242.382	27.440	42.412

Notes: Table E2 reports the heterogeneous impact of HSR openings on the scale of enterprises. ***p< 0.01, **p <0.05, *p < 0.1. Robust standard errors are clustered at the prefecture levels. Column (1) represents the total scale of large industrial enterprises, and columns (2)-(4) represent the total scale of domestic, Hong Kong, Macao, Taiwan, and foreign-funded enterprises, 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 The heterogeneous impact of HSR opening on trust and evaluation

Dependent variable	(1)	(2) *Political trust*	(3)	(4)	(5) *High Trust*	(6)
	All	Locals	Migrants	All	Locals	Migrants
HSR	0.342***	0.377***	-2.500*	0.062***	0.066***	-0.262
	(0.113)	(0.113)	(1.435)	(0.021)	(0.021)	(0.263)
HSR×*lndistance*	-0.055**	-0.063***	0.410	-0.010**	-0.011***	0.027
	(0.023)	(0.023)	(0.309)	(0.004)	(0.004)	(0.057)
Individual Controls	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓
Observations	79,071	77,565	1,495	79,071	77,565	1,495
R-squared	0.070	0.071	0.161	0.055	0.055	0.151
Outcome mean	5.027	5.035	4.572	0.367	0.368	0.300

Notes: Table E3 reports the heterogeneous impact of HSR opening on political trust and government evaluation. The data is sourced from individual-year level data from the 2012-2018 China Family Panel Studies (CFPS). Robust standard errors are clustered at the prefecture levels. The dependent variables in columns (1)-(3) are the respondents' political trust in local government officials (*Political trust*), while those in columns (4)-(6) are a dummy indicating high trust levels (*High Trust*). *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 E4 HSR opening, discrimination, and protest.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dialect diversity		Predetermined floating population ratio		Predetermined economic disparities		Hukou discrimination gap	
	High	Low	High	Low	High	Low	High	Low
HSR	0.534***	0.373	0.544***	0.405**	0.505**	0.435***	0.506**	0.360
	(0.152)	(0.230)	(0.147)	(0.197)	(0.244)	(0.157)	(0.215)	(0.254)
HSR×*lndistance*	-0.119***	-0.093**	-0.132***	-0.093**	-0.114**	-0.106***	-0.112**	-0.099*
	(0.031)	(0.047)	(0.035)	(0.038)	(0.051)	(0.032)	(0.043)	(0.054)
Predetermined-year Control	✓	✓	✓	✓	✓	✓	✓	✓
Prefecture FE	✓	✓	✓	✓	✓	✓	✓	✓
Province-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	888	1,008	1,008	888	973	923	1,406	476
R-squared	0.907	0.905	0.903	0.903	0.921	0.903	0.893	0.956

Notes: Table E4 reports the heterogeneous treatment effects by discrimination, where we employ four different measures to shed light on the discrimination in provincial capitals against migrant workers. ***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. Columns (1) and (2) group by the average number of distinct dialect areas per prefecture at the province level. Columns (3) and (4) group by the proportion of floating population in the provincial capitals in 2010. Columns (5) and (6) group by the per capita GDP ratio between provincial capitals and regions farther from them (greater than the median distance) in 2010. Columns (7) and (8) group by the difference in perceived hukou discrimination between the aggregated rural and urban hukou groups, based on the CSS dataset. "High" refers to values above or equal to the median, and "Low" refers to values below the median.

TABLE E5 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 E5 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.

TABLE E6 Alternative explanations: HSR-related portest.

	(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 E6 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. The dependent variables in columns (1) and (4) are the standard number of protests and real number of protests, respectively. Columns (2) and (5) exclude those related to HSR opening based on columns (1) and (4), whereas columns (3) and (6) focus only on HSR-related protests. In column (7), we return to the standard number of protests, but code HSR opening as non-absorbing (*Instant HSR*). *HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access and 0 otherwise. *Instant HSR* is a dummy variable that takes the value of 1 if a prefecture has HSR access this year, and 0 otherwise. The variable *lndistance* represents the distance (logged) between the prefecture and its provincial capital.

TABLE E7 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 E7 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. *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.