Information Super-Railway?

Transportation Infrastructure and Information Production in Financial Markets*

Kejing Chen[1] J. Daniel Chi[2] Wenqi Guo[3]

Abstract

When large numbers of people are enabled to travel rapidly around the country, will it affect information production in financial markets? We study a government-directed buildout of high-speed rails (HSR) and find that after being connected to the HSR network, firms experience increased information-production activities from market participants. Stock prices, the aggregator of all information, become more liquid and informative and less prone to crashes. These HSR-induced information effects are independent from the effects of firm performance or economic development. A severe HSR accident caused by a lightning strike allows us to observe the counterfactuals and further confirm causality.

Keywords: transportation infrastructure, information, financial markets, high-speed rail

JEL codes: G14, G18, H54, O18, R40

---

* For helpful comments and suggestions, we thank Zhi Da, Jesse Ellis, Kristine Hankins, Jiekun Huang, Jerry Hoberg, Matthew Jaremski, Yanxi Li, Randall Morck, Ankur Pareek, Kelly Shue, Xunhua Su, Mike Sullivan, Baolian Wang, Xiong Xiong, Liyan Yang, Zihan Ye, Andrew Zhang, seminar participants at DUFE and UNLV, and conference participants at FMA. Kejing Chen acknowledges support from the National Natural Science Foundation of China (grant numbers: 71802043 and 71790594). Daniel Chi acknowledges research grant support from the Troesh Center for Entrepreneurship and Innovation and the Rocky Mountain Center for International Business Education and Research. This research was developed in part under grants from the U.S. Department of Education through its Centers for International Business Education and Research at University of Colorado Denver and Brigham Young University. This research does not necessarily represent the policies of the U.S. Department of Education.

[1] Dongbei University of Finance and Economics, China; chenkejing@dufe.edu.cn.
[2] Corresponding author. Department of Finance, University of Nevada, Las Vegas; 4505 S. Maryland Parkway, Las Vegas, NV 89154-6008, USA; daniel.chi@unlv.edu.
[3] Dongbei University of Finance and Economics, China; gracewguo@163.com.

"*You know, Chris? Do you know that it takes 4 hours and 18 minutes to take the bullet train from Beijing to Shanghai, and it takes 21 hours to take the train from New York to Chicago? And they're both about the same distance.*"

– Thomas Friedman to Chris Cuomo on CNN's *Cuomo Prime Time*, February 8, 2021

"*This law makes the most significant investment in passenger rail in the past 50 years... Next year will be the first year in 20 years American infrastructure investment will grow faster than China's.*"

– Joe Biden at the signing ceremony of the Infrastructure Investment and Jobs Act, November 15, 2021[1]

1. Introduction

Transportation infrastructure plays a vital role in the economy (e.g., Donaldson and Hornbeck, 2016; Asher and Novosad, 2020), yet little is known about its effects on financial markets. This is surprising given financial markets' prominence in the economy: the financial industry has become the largest economic sector; financial-assets-to-GDP has risen to a record level; and "financialization" has been prevalent around the world.[2] On the other hand, however, it is not immediately obvious that transportation infrastructure affects financial markets: capital rarely travels on roads or railways anymore. Information, which directs capital flow and is often called the most valuable commodity,[3] can now also travel electronically. So, to an investor in New York City who is considering an investment opportunity in Chicago, does it matter whether the train takes 4 hours or 21 hours to get to Chicago? Put in broader terms, does transportation infrastructure affect financial markets?

This research question becomes even more relevant in light of the current debate on infrastructure investment. A recent global survey confirms the prevailing view that infra-

[1] The transcripts for the CNN interview and Biden's speech can be found here and here.
[2] GDP-by-industry data can be found at BEA. Various ratios that measure financial-assets/GDP can be calculated using data from FRED. Krippner (2005) documents the financialization of the U.S. economy. Van der Zwan (2014) surveys the financialization literature.
[3] This notion does not have an exact origin but is certainly popularized by the movie Wall Street (1987), in which Gordon Gekko (by Michael Douglas) said "The most valuable commodity I know of is information."

structure investment is important to the economy. However, only 30% of U.S. respondents are satisfied with existing infrastructure, which ranks toward the bottom among the surveyed countries.[4] The wide gap between desire and reality has brought infrastructure investment to the forefront of policy debate.[5] Given that infrastructure and financial markets are essential components of the economy, it seems necessary to explore the relation between the two.

We hypothesize that transportation infrastructure plays a vital role for financial markets by facilitating information production. While hard information can be quantified and transmitted electronically, soft information is contextual, qualitative, and difficult to verify (Liberti and Petersen, 2019). As Albert Einstein once said, not everything that counts can be counted. Soft information certainly counts but is difficult to be counted. Soft information, in Hayek's words, is "*the knowledge of the particular circumstances of time and place*" (1945, p. 521). Transportation infrastructure allows people to access those *particular circumstances of time and place* and to comprehend and verify information in its context, and therefore facilitates information production.

To test this hypothesis, we employ the staggered buildout of high-speed rails (HSR) in China as a transportation-infrastructure shock to firms along the HSR lines. The Chinese government directed this rapid and massive buildout: from no HSR in 2007, China had built 35,000 kilometers of HSR lines by the end of 2019. China's HSR network accounts for about 70% of the world's total and is by far the largest passenger-dedicated HSR network in the world. In the words of Martin Raiser, World Bank Country Director for China, "*Large numbers of people are now able to travel more easily and reliably than ever before.*"[6]

We implement a generalized difference-in-differences (DiD) identification strategy to isolate the treatment effect of HSR connections. We find that after a firm is connected to

[4] See the 2020 Ipsos survey for more details. The 30% satisfaction rate is below the average of 43% across the 27 countries in the survey, and the 90% satisfaction rate in China.

[5] In his presidential campaign, Joe Biden outlined an ambitious infrastructure plan, in which he proposed an "*end-to-end high-speed rail system that will connect the coasts*" and "*spark the second great railroad revolution.*" On November 15, 2021, he signed the largest infrastructure investment bill in recent history.

[6] From World Bank Press Release of July 8, 2019. The HSR network transports only people, not products.

HSR, information-production activities about the firm increase significantly. Institutional investors, the media, and analysts visit the firm more often. Institutions increase their equity ownership, the media produce more news stories, and analysts increase their coverage. Stock prices, the aggregator of all information, reflect the increased information production: liquidity increases, crash risk decreases, and stock prices reflect more firm-specific information and move less synchronously with the market. Controlling for an extensive set of firm-performance or economic-growth measures barely changes the estimated effects of HSR connections. Heterogeneity analysis shows that the HSR effect is driven by firms that face relatively poor information environments, such as those with higher levels of unexplained accounting accruals, employing less reputable auditors, or located in areas with low Internet penetration. The collective evidence suggests that transportation infrastructure facilitates information production, and that the information effect is independent from the effects of firm performance or economic growth.

Studying the setting of the HSR buildout offers several advantages. First, during earlier infrastructure buildouts, such as railways or the U.S. Interstate Highway System, much information still travelled on paper through the transportation infrastructure. In contrast, information today can travel digitally. Therefore, even if transportation infrastructure facilitated information production in the past, it may not today. The HSR buildout is recent and still ongoing, so our findings have external validity in the current age.

Second, rather than studying an infrastructure project that is initiated in response to profitable situations and thus endogenous to economic development, we need to study a project that is exogenous so that we can draw causal inferences. Of course, as Robert Shiller (2017, p. 968) once quipped, "*almost nothing beyond spots on the sun is truly exogenous in economics.*" The moral is that although economic shocks are rarely random or "truly" exogenous, we can still draw causal inferences from non-random but *primarily* exogenous shocks, such as passages of laws or regulations (e.g., Bertrand and Mullainathan, 2003; Duarte, Han, Harford, and Young, 2008; and Faccio and Xu, 2015). In our setting, HSR connections are primarily exogenous to firms along HSR lines because these firms had no input into the decision making of HSR connections – just like how law or regulation changes, although not random, can be primarily exogenous to individual firms. Of course, omitted variables may simultaneously drive HSR connections and information production. We address this concern extensively in the empirical tests.

Third, the HSR network transports only people, not products. This passenger-dedicated feature helps us better identify HSR's information effect, which stems mostly from the movement of people.

Since HSR connections are not randomly assigned, we cannot observe the counterfactuals. That is, for firms that are connected to HSR, we cannot observe what would have happened if they were not connected to HSR (or what will happen if they are disconnected from HSR). A severe HSR accident provides an exogenous shock that allows us to at least partially observe the counterfactuals of HSR connections.

On July 23, 2011, lightning struck a section of HSR's power network in Zhejiang Province and caused a massive control-system malfunction. As a result, a fast-moving train rear-ended into a slow-moving train, killing 40 passengers. The consequences of this accident were wide-ranging and led to a nationwide reduction in HSR speed. The reduced speed diminished the advantage of HSR relative to other forms of transportation. More importantly, the accident shattered the public's confidence in HSR's safety and deterred many people from riding HSR. Passenger-per-km, after increasing by 51% from 2009 to 2010, dropped by 4% from 2011 to 2012. We find that in the aftermath of the accident, HSR's effect on information production was significantly reversed. This lends support to the causal interpretation of the HSR effect. In addition, we find no evidence that the accident caused significant economic damage as measured by GDP. Thus, the reversal in information production cannot be attributed to impaired economic growth. This is additional evidence that the information effect of HSR is independent from the effect of economic growth.

We conduct additional tests to further address various endogeneity concerns. First, we find that the increase in information production occurs only after HSR connections, not before, suggesting that causality runs from HSR to information production rather than the other way around. Second, it is possible that because HSR lines are intended to run through large cities, which also have better information environments, the positive relation between HSR and information production might be driven by large cities. We exclude large cities and focus on smaller cities, which are "inconsequential" in the sense that these cities just happen to be the HSR routes, and their economic development has little bearing on the decision of HSR connections.[7] The HSR effect is stronger in these smaller cities, again consistent with causality running from HSR to information production.

Another endogeneity concern is that correlations in underlying economic trends may have produced the findings. Specifically, China's economic and financial development, and hence information production, have been trending upwards. Meanwhile, more firms have been connected to HSR over time. Is it probable that these underlying trends have led to the positive relation between HSR connections and information production? We first note that our DiD tests satisfy the parallel-trend requirement and already account for such effects. We further conduct falsification tests by running simulations to randomly select a city or a firm for HSR connection. If the baseline findings are the results of correlations in underlying trends, the falsification tests should return significant results. But they do not.

We also employ two instrumental variables for HSR connections: a city's terrain gradient and whether the city was connected to a railway a half century ago. A less hilly terrain and the existence of conventional railway tracks make HSR construction easier.[8] We find that gradient and historical railway connections are strong predictors of HSR connections, and the instrumented HSR has a positive effect on information production. Collectively, the empirical evidence supports our hypothesis that HSR connections enhance information production.

The importance of information production has been well recognized since at least Akerlof (1970) and Hölmstrom (1979) and is a main topic in finance research (e.g., Gao and Huang, 2020). Extensive efforts have been devoted to identify specific information-

[7] For example, see Chandra and Thompson (2000) and Michaels (2008) for the "inconsequential places" approach.
[8] The "historical route IV" approach was developed by Duranton and Turner (2012) and adopted by later studies.

production channels such as institutional investors (Shleifer and Vishny, 1986), the media (Engelberg and Parsons, 2011), analysts (Chen, Harford, and Lin, 2015), and investor protection (Morck, Yeung, and Yu, 2000). While hard information can be easily quantified and transmitted electronically, the production of soft information requires extensive human involvement. We show that even in the age of the Internet, transportation infrastructure, by facilitating the travel of people, encourages the production of soft information and is clearly relevant for financial markets.

Our study complements previous studies on geographical proximity and information production (e.g., Petersen and Rajan, 2002).[9] One challenge for these studies is that geographical proximity is the outcome of self-selection and is thus endogenous to many aspects of financial markets. Several studies use the introduction of new airline routes to isolate the treatment effect of proximity (e.g., Giroud, 2013; Ellis et al., 2020; and Da et al., 2021), yet initiations of new routes are driven by airlines' profit motive, which inevitably responds to local economic development. Airline connections also affect relatively few people. We study a transportation-infrastructure project that is of a much larger scale. The project is government-driven and primarily exogenous to local economic development and individual firms. As such, our findings provide direct evidence on how transportation infrastructure can affect financial markets.

Our study extends the broad literature on how transportation infrastructure affects economic activities. While there is evidence on how transportation infrastructure affects various aspects of the economy (e.g., Baum-Snow, 2007; Redding and Turner, 2015; Donaldson and Hornbeck, 2016; Donaldson, 2018; Asher and Novosad, 2020), little is known about its effects on financial markets. Highways, airports, and conventional railways move people as well as products. In comparison, HSR only moves people, and the movement of people engenders the movement of information. Therefore, we are able to more sharply identify the information effect of transportation infrastructure and its impacts on financial

[9] Other examples of proximity studies are Loughran and Schultz (2005), John et al. (2011), and Coval and Moskowitz (1999, 2001).

markets.[10] Furthermore, we show that the information effect is independent from the effects of economic development, firm performance, or investment opportunities.

2. Research Questions and Identification Strategy

2.1. Background of China's HSR

Several characteristics of the HSR buildout make it a powerful setting to answer our research questions. First, the buildout was centrally planned and macro-policy driven. The plan started with a 4×4 grid – representing 4 north-south routes and 4 east-west routes – and was later expanded to an 8×8 grid.[11] This reduces the concern that some local economic shocks reverse-caused the connection of HSR to a locality.[12] A quote from Lawrence et al. (2019, p. 2) adds more color to the centrally-planned nature of the buildout:

> *"The first chapter outlines … the key role played by the Medium- and Long-Term Railway Plan (MLTRP). This plan, first approved in 2004 with revisions in 2008 and 2016, looks up to 15 years ahead and is complemented by a series of Five-Year Plans, prepared as part of the general planning cycle. These plans are rarely changed once approved."* [emphasis added]

One result of the central planning is the striking speed of the buildout, which makes the treatment more exogenous to local firms. Starting in 2008 with only 672 kilometers of HSR lines, the operating mileage grew at an average annual rate of 43% and reached 35,000 kilometers by the end of 2019, accounting for about 70% of the world's total. Figure 1 provides a color-coded map that shows the progression of the HSR buildout.

[Figure 1 here.]

[10] Several studies examine the economic effects of infrastructure projects in China (e.g., Faber, 2014; Baum-Snow et al., 2017; Banerjee et al., 2020; Kuang et al., 2021; and Xiong et al., 2021), but they do not study the broad impacts of HSR on information production in financial markets, which is the focus of our study.
[11] A recent and comprehensive information source on China's HSR development is a World Bank study by Lawrence, Bullock, and Liu (2019), titled *China's High-Speed Rail Development*. Most of our HSR-related statistics are from this study.
[12] Indeed, there are plenty criticisms that infrastructure projects in China are often inefficient and wasteful because they are not guided by economic motives. For example, see the academic study by Ansar et al. (2016) as well as numerous news stories, e.g., China's empty airports (by Reuters) and ghost towns (by CNN).

Second, as mentioned earlier, HSR transports only people, not products. This passenger-dedicated feature helps us more sharply identify HSR's information effect, which stems mostly from the movement of people. The movement of people engenders the movement of information, especially the kind of information that is not easily quantified or electronically transmitted.

Third, China is a vast country, and HSR significantly reduces travel time. For example, HSR trains travel at up to 350 km/hour, several times faster than conventional trains. The HSR connection between Beijing and Shanghai reduced travel time from 13 hours to four and a half hours. As of 2019, there were about 250 pairs of HSR trains running on the Beijing-Shanghai line *each day*, and 44 ran the entire distance.

Fourth, the HSR network has not only markedly reduced the travel time but also significantly expanded travel capacity. By 2019, the HSR had become a ubiquitous part of life in China: there were over 2,600 pairs of HSR trains each day; most HSR lines had at least an hourly service between 7:00 a.m. and midnight; and the HSR network carried 56% of the 8.3 million passengers that used China's nonurban rail network each day. Surveys have shown that for trips between 150 and 800 kilometers in distance, travelling by HSR predominates over travelling by highway or air. In addition to taking market shares from other transportation forms, HSR generates 10-20% new trips that otherwise would not have occurred. About half of all HSR trips are for business purposes.

Another rather unique feature of China's transportation infrastructure is the notoriously low punctuality of air travel. A BBC article in 2016 vividly documents flight delays and mass cancellations in China.[13] For example, in 2015 China's airports were among the least punctual in the world: the on-time rating for Shanghai's Pudong Airport was 52%; for Beijing, 64%. In comparison, Atlanta International Airport, the busiest airport in the U.S., had an on-time rating of 84%; Tokyo's Haneda Airport, the busiest airport in Japan, 92%. The article identifies a major cause for China's dismal on-time ratings: the airspace is mostly controlled by the military – "*less than 30% of China's airspace can be used by commercial*

[13] "This is why China's airports are a nightmare," by Justin Bergman, April 28, 2016.

airlines, compared to about 80% of the airspace in the United States." The article goes on to highlight the punctuality of China's high-speed rail, which consistently has an on-time rating close to 100%, and that "*many businesspeople have stopped flying altogether on short routes and rely on the high-speed train network*."

In summary, the HSR buildout has been a powerful exogenous shock and has reduced travel time for many people, including for businesses and their investors. Reduced travel time and increased travel convenience reduce frictions in information gathering, and should enhance information discovery and dissemination.

2.2. Research Questions

Our main hypothesis is that HSR connections enhance information production, especially the "soft" kind of information that is not easily quantified and electronically transmitted. To test this hypothesis, we conduct two sets of tests.

The first is information production by external agents, such as institutional investors, the media, and stock analysts. They all have strong motivation to collect soft information, understand the context of the information, and disseminate the information in its context. For example, institutional investors strive to generate positive alpha, the media strive to write news with interesting context, and analysts strive to provide unique insights to help their clients generate alpha. All these goals require these agents to go beyond the hard information that is readily available to everyone and to acquire the soft information that will set them apart from the competition. Firms listed on the Shenzhen Stock Exchange are required to disclose when investors, the media, or analysts visit the firm. We predict that after a firm is connected to HSR, these external agents will pay greater attention to the firm, such as visiting the firm more often, increasing investment in the firm, reporting more news stories, or producing more research reports.

Other than the information channels discussed above, HSR connections likely work through additional information channels that are difficult to quantify or not captured by available data. Regardless, stock price is the aggregator of all information, and increased information production, whether through observable or unobservable channels, should lead to greater stock liquidity as well as more informative stock prices. That is, stock price movements should be driven more by firm-specific information relative to marketwide information, resulting in lower synchronicity with the market. Jin and Myers (2006) show that to protect their rent extraction, insiders are more likely to conceal negative information than positive information. Kothari et al. (2009) find that managers indeed delay the disclosure of negative information. Over time, negative information accumulates to a breaking point and is released all at once, causing stock crashes. Thus, increased information production should also reduce stock crash risk. Our second set of empirical tests is to examine whether HSR connections increase stock liquidity, reduce crash risk, and reduce stock price synchronicity.

2.3. Identification Strategy

To estimate the treatment effect of HSR connections on information production, we implement a difference-in-differences (DiD) identification strategy. Firms that have been connected to HSR are in the treatment group. Firms that have not been connected (yet) are in the control group. Because HSR buildout is staggered, a firm remains in the control group until it is treated – some firms stay in the control group for the entire sample period. An example will help illustrate the DiD design. Changhong Electric, a leading TV manufacturer in China, is headquartered in the city of Mianyang in southwest China.[14] It was connected to an HSR line in 2014. We first measure the difference in information production of Changhong Electric before and after 2014. However, other things in 2014, such as overall economic development, may also have affected information production. Therefore, we also measure the difference in information production of the control group before and after 2014. We then compare the two differences to estimate the HSR treatment effect.

The test is implemented through a generalized DiD regression model (e.g., Bertrand and Mullainathan, 2003):

[14] Changhong became publicly traded in 1994 and is listed on the Shanghai Stock Exchange. In 2004, 90% of the TV sets that the U.S. imported from China were made by Changhong.

$$InfoProd_{i,t} = \alpha_0 + \beta\ PostHSR_{i,t} + \gamma X_{i,t} + \mu_i + \mu_t + \varepsilon_{i,t}, \tag{1}$$

where $InfoProd_{i,t}$ is one of the information-production measures for firm i in year t, and $PostHSR_{i,t}$ is an indicator variable equal to 1 if firm i's headquarters city has been connected to HSR by year t. X is a set of variables that control for firm and local characteristics that may affect both information production and HSR connections, such as local economic development. μ_i are firm fixed effects that control for fixed differences between the treatment and the control groups. μ_t are year fixed effects that control for aggregate fluctuations over time. α_0 is the common intercept, and $\varepsilon_{i,t}$ is the remaining residual. β measures the treatment effect of HSR connections on information production and is the main coefficient of interest. We predict β to be of the sign that indicates increased information production. We supplement this DiD identification strategy with a battery of additional tests, which we will explain in the corresponding sections.

3. Data

3.1. Data Sources

We start the data set with all domestically listed firms in China between 2007 and 2018, and collect financial data from the China Stock Market and Accounting Research (CSMAR) database. We then exclude the following firms: firms at risk of being delisted from exchanges; financial firms; firms that are listed within 1 year; firms that have annual trading data fewer than 30 weeks; and firms that do not have the requisite control variables. The final data set has 25,600 firm-year observations.

Next, from the official websites of China Railway State Group, we use web-crawling technology to collect the dates that each city was connected to an HSR line.[15] We manually clean the data and construct a set of HSR connection dates for each city. For cities that are connected to more than one HSR line, we use the date of the first connection. We match this data set with the financial data set from CSMAR, and apply the HSR connection date of each firm's headquarters city as the connection date of the firm.

[15] http://www.china-railway.com.cn and https://www.12306.cn.

Besides CSMAR, we also collect financial data from Wind Economic Database. Data on institutional investors, media coverage, and stock analysts are from Chinese Research Data Services Platform (CNRDS). The media database covers over 400 news websites and over 600 newspapers. We count only original news stories, not redistributed stories. Geographic data, which we use to construct the instrumental variable for *PostHSR*, are from the Geospatial Data Cloud by China Academy of Sciences.[16] We use ArcGIS software to construct an elevation map. For each 200-meter-by-200-meter grid on the map, we compute the elevation difference between the highest point and the lowest point. The average of these elevation differences for all grids within a city's boundary is the city's *gradient*, which measures the city's hilliness. Appendix A lists the detailed data source for each variable.

3.2. Measures of Information Production

As shown in Equation 1, the dependent variable is the set of information-production measures at the firm-year level. The first set of information-production measures are those of market participants:

- Institutional investor activities (the number of site visits by institutional investors, the number of institutional investors, and the percentage of firm equity held by institutional investors).
- Media activities (the number of site visits by the media, the number of Internet news stories covering the firm, and the number of newspaper stories covering the firm).
- Analyst activities (the number of site visits to the firm by analysts, the number of analysts following the firm, and the number of analyst research reports covering the firm).

The site-visit data are only available for firms listed on the Shenzhen Stock Exchange, so for the site-visit tests, we limit the sample to Shenzhen-listed firms.

[16] http://www.gscloud.cn/

The second set of information-production measures are related to stock prices:

- Stock liquidity as measured by 1) the Amihud (2002) illiquidity measure and 2) the bid-ask spread of stock prices. Lower values indicate higher liquidity.
- Stock crash risk as measured by 1) the negative conditional skewness of weekly stock returns and 2) the down-to-up volatility of weekly stock returns (e.g., see Chen, Hong, and Stein, 2001). Higher values indicate higher crash risk.
- Stock-price synchronicity that captures the composite effect of information production.

Following the synchronicity literature, we run the following annual regression for each firm, using weekly stock returns, to obtain R^2.

$$R_{i,w,t} = \alpha + \beta_1 R_{M,w,t} + \beta_2 R_{M,w-1,t} + \beta_3 R_{I,w,t} + \beta_4 R_{I,w-1,t} + \varepsilon_{i,w,t}, \tag{2}$$

where $R_{i,w,t}$ is firm i's week w stock return (including dividends) in year t, R_M are the value-weighted market returns for weeks w and w-1, and R_I are the value-weighted industry returns (excluding firm i itself from the industry). Industry classification is based on the guidelines by China Securities Regulatory Commission. Because a regression R^2 is bounded between 0 and 1, we follow the literature and calculate synchronicity as follows.

$$SYNCH_{i,t} = log\left[{R_{i,t}}^2/\left(1 - {R_{i,t}}^2\right)\right] \tag{3}$$

Lower R^2, and hence lower *SYNCH*, indicates lower synchronicity of stock prices. That is, stock prices contain more firm-specific information relative to market-level information.[17]

[17] There is an argument that lower R^2 may reflect less informative, rather than more informative, stock prices (e.g., West, 1988; Dasgupta et al., 2010; Chan and Chan, 2014). On the other hand, there are numerous studies suggesting that lower R^2 reflect more informative stock prices (e.g., Morck et al., 2000; Durnev et al., 2004; Jin and Myers, 2006; Hutton et al., 2009; Gul et al., 2010; and Gao and Huang, 2020). We note that empirical evidence that suggests "low R^2 means less information" is often based on specific events, such as seasoned equity offerings (SEOs) and cross-listings. In comparison, empirical evidence that suggests "low R^2 means more information" is mostly based on large samples rather than specific events. While specific events may increase the test power, the interpretation is often challenged by potential self-section bias. For example, SEOs and cross-listings are self-selected activities, and those firms' information environment around those events may not be representative of the cross section of firms. To reconcile these studies is beyond the scope of this paper. We do note that when trying to interpret empirical findings, it is imperative to understand the nature of the information-production activities being studied. In this respect, based on the information-production activities that we study, the findings suggest that the reduction in R^2 induced by HSR connections is consist with increased, rather than decreased, information production.

3.3. Control Variables

As mentioned earlier, the key endogeneity concern is that some omitted variables are driving both information production and HSR connections. The primary candidate for this omitted variable is local economic growth, which can arguably enhance information production and necessitate the connection of HSR to the locality. To mitigate this concern, we control for a large set of firm and city characteristics that might be related to both information production and *PostHSR*. The firm-level controls include firm size and age, profitability, growth rate, market-to-book ratio, R&D expense, financial leverage, institutional holdings, whether using a big auditing firm, whether is a state-owned enterprise, and stock trading volume. The city-level controls include per-capita GDP, foreign investment, per-capita road area, retail sales, and population size. In addition, we control for unobservable firm and year heterogeneity by including firm and year fixed effects.

Detailed variable definitions are in Appendix A. When performing various robustness checks, we employ additional variables, which we will describe in those sections.

3.4. Summary Statistics

Panel A of Table 1 presents summary statistics of the regression variables. To mitigate the influence of extreme values, we winsorize all continuous variables at the 1st and 99th percentiles of the sample. Starting from the top of the panel, we see that each year, a typical firm receives about 10 site visits by institutional investors and is 37% owned by an average of 38 institutional investors; it receives fewer than 1 site visit by the media, and is covered by 266 Internet news stories and 73 newspaper news stories; it receives about 17 analyst site visits, is followed by 7 analysts, and is covered by 15 analyst reports. A typical stock's returns are negatively skewed and have greater downside volatility. For R^2, both the mean and the median are around 48%. The standard deviations of these variables are sufficiently large.

PostHSR has a mean of 0.73 and a median of 1. That is, most firms are treated during the sample period. The standard deviation is 0.45, which provides a sufficiently large dispersion and enables us to identify the treatment effect. Statistics for the control variables look reasonable and are consistent with other studies that employ financial data of Chinese listed companies. For example, the average M/B ratio is 2.1, the average ROE is about 7%, and the financial leverage is about 44%.

[Table 1 here.]

Panel B provides a detailed chronological account of the HSR treatment. As expected, progressively more firms and cities get treated over time. By 2018, 95% of firms and 77% of cities in the sample are treated. This lack of dispersion in the treatment variable in later years actually reduces the test power and biases against our finding significant treatment effects. In a robustness check, we limit the sample to up to 2014 and obtain similar results.

4. Empirical Findings

4.1. Univariate Results

Table 2 presents correlation coefficients of the regression variables – below the diagonal line are Pearson correlation coefficients; above the line are Spearman's rank correlation coefficients. Without controlling for any other factor, *PostHSR* is positively correlated with information production measures of external agents: institutional investors, the media, and analysts. Examining financial-markets measures, *PostHSR* is negatively correlated with the two illiquidity measures and the two crash-risk measures, as well as with *SYNCH*, meaning that HSR connections are correlated with higher liquidity, lower crash risk, and less synchronous (i.e., more idiosyncratic) stock price movement. The signs of all these correlations are consistent with our prediction that HSR connections are associated with greater information production.

[Table 2 here.]

In Table 3, we perform a univariate test on the key variables between the treatment group and the control group. Consistent with our prediction and the results in Table 2, the comparison shows that information production in the treatment group is significantly greater than the control group: activities by institutions, the media, and analysts are all greater; stock liquidity is higher; crash risk is lower; and R^2 and *SYNCH* are lower. The difference in *SYNCH* is 0.088, about 25% of its standard deviation, and is economically large.

The two groups are also different in various other dimensions. For example, the treatment group is slightly larger, spends more on R&D, has lower financial leverage, is more likely to be audited by a big auditor, is less likely to be an SOE, and has a lower share turnover. Therefore, it is imperative that we control for these various dimensions in the multivariate tests, which we do in the next section.

[Table 3 here.]

4.2. Information Production by Market Participants

Institutional investors are often large shareholders and have a strong incentive to discover firm-specific information (e.g., Shleifer and Vishny, 1986). Given that HSR connections significantly reduce travel time and increase travel convenience, do institutions visit firms more often and increase their investment? We use three measures of institutional investor activities for each firm-year observation: the number of institutional investor site visits, the percentage of institutional stock holdings, and the number of institutional investors. We implement our main DiD identification strategy of Equation 1 and show the results in Table 4. The coefficient estimates on *PostHSR* are positive and significant, suggesting that HSR connections have a significantly positive effect on institutional investor activities: the number of site visits increases by 1.9, institutional holdings increase by 1.2 percentage points, and the number of institutional investors increases by 1.6.[18]

Examining the control variables, we see, unsurprisingly, that larger, older, more profitable firms, as well as firms with larger growth opportunities (as proxied by higher M/B and R&D), attract greater institutional-investor attention. Fast-growing firms are held by fewer institutions but receive more site visits, presumably because their recent growth

[18] The *t*-statistics are based on standard errors adjusted for firm-level clustering. In untabulated tests, we adjust for both firm-level clustering and year clustering, and the *t*-stats barely change. Petersen (2009) suggests that for panel data sets with a large cross section and a relatively short time span, which is the case for our data set, including firm and year fixed effects and adjust for firm-level clustering is sufficient to correct the bias in standard errors that is caused by clustered observations.

attracts increased attention. Local GDP is positively related with institutional attention. Other city-level characteristics are less significant, likely because firm-level characteristics already subsume the explanatory power of these city-level characteristics.

As already mentioned, the main endogeneity concern is that rather than HSR connections enhancing information production, it is some other variable, such as a locality's economic development, that affects both information production and the connection of HSR.[19] If so, we should observe the changes in information production before the treatment of HSR connections. To test whether this is the case, we implement a dynamic regression by replacing *PostHSR* with a series of pre- and post-HSR indicators. Specifically, for each of the three years before and two years after an HSR connection, we construct an indicator variable that equals one for that year. We want to test whether the HSR effect exists before an HSR connection. We plot the coefficient estimates of these pre- and post-connection indicators in Figure 2. There is no treatment effect before HSR connections. This also shows that our DiD specification satisfies the parallel-trend requirement. In contrast, there is significant treatment effect for the connection year and two years after. This test provides additional evidence that the direction of causality is from HSR connections to information production.

[Table 4 & Figure 2 here.]

Information production is a main function of the media, and media coverage impacts financial markets (e.g., Fang and Peress, 2009; Engelberg and Parsons, 2011). After a firm is connected to an HSR line, does the firm attract greater media attention? For each firm, we measure media attention by the number of site visits by the media, the number of Internet news stories covering the firm, and the number of newspaper stories covering the firm. We then implement tests that are parallel to those for institutional investors. The results are in Table 5. HSR connections have a significantly positive effect on media attention. The number of site visits by the media increases – the statistical significance is

[19] A potential way to control for time-varying local economic development is to include city-by-year fixed effects in the regressions. Unfortunately, we cannot implement this because *PostHSR* is measured at the city-year level and will be perfectly collinear with the city-by-year fixed effects.

high, but the economic magnitude seems small. However, given that the average number of media site visits is only 0.119 and the median is 0, the coefficient of 0.038 is economically meaningful.

The seemingly low level of mean and median site visits is probably not that surprising considering the vast number of firms that each news agency covers. Besides, the media often travel to the vicinity of a firm and interview people about the firm. This is what we often see on TV, where a reporter stands in front of the logo of a firm's building and interviews many people, but not the firm's management. In a way, these outsiders' views are what give context to a news story and make it interesting. Also, since HSR allows more people to get geographically close to a firm, a reporter can interview those people without physically travelling to the firm. Therefore, HSR's effect on media activities should ultimately be reflected in a greater number of news stories. Test results in Columns 2 and 3 are consistent with this prediction. After a firm is connected to HSR, stories by Internet news outlets increase by 23.2 per year, while stories by traditional news outlets increase by 8.5 per year. We again conduct the parallel-trend test and plot the result in Figure 3. There is no HSR effect before connections, but significant effects occur post-connections.

[Table 5 & Figure 3 here.]

It is well documented that stock analysts play an instrumental role in discovering and disseminating firm-specific information (e.g., Asquith et al., 2005). A significant part of analysts' insights comes from private meetings with management and visiting companies to discover soft information (e.g., Soltes, 2014; Cheng et al., 2016). We employ three measures of analyst activities – the number of analyst site visits, the number of analyst following, and the number of analyst reports. The tests are parallel to those on institutional investors and the media. The results, in Table 6, show that HSR connections increase analyst activities. The coefficient estimates suggest that after HSR connections, the number of analyst site visits increases by 3.8 per year, the number of analysts increases by 0.6, and the number of analyst reports increases by 2.2 per year. We again conduct the parallel-trend test and plot the result in Figure 4. There is no HSR effect before connections, but significant effects occur post-connections.

[Table 6 & Figure 4 here.]

In untabulated tests, we find that analyst forecast precision increases after firms are connected to HSR. Three contemporaneous papers find similar results (Kong et al., 2020; Chen et al., 2021; and Chen et al., 2021). We extend these studies by showing a wide range of impacts of HSR on information production and the effects on stock prices.

Collectively, evidence in this section provides a coherent picture: HSR connections bring about increased institutional ownership, increased media coverage, and increased analyst following. All three channels contribute to enhanced information production.[20]

4.3. Stock Liquidity

Since HSR connections lead to increased information production by market participants, the increased information should be incorporated into stock prices. First, we expect that increased information production should reduce information asymmetry and lead to higher stock liquidity. The results in Table 7 indeed show that HSR connections significantly reduce the price impact of stock trading (as measured by the Amihud (2002) illiquidity) as well as the bid-ask spread. That is, HSR connections significantly increase stock liquidity. The coefficient estimates suggest that the illiquidity measure and the bid-ask spread are respectively lower by 4.5% and 7.6% of one standard deviation.[21] The parallel-trend test,

[20] In an untabulated test, we identify the year when a firm's city and an analyst's city were connected by a direct HSR line. We use this direct-connection year to define an alternative HSR-treatment variable for tests on analyst activities. Since each firm can have multiple analysts, the sample transformed from two dimensions (firm-year) to three dimensions (firm-year-analyst_city). Similarly, we define alternative HSR-treatment variables for tests on institutional investors and the media. Results using these alternative treatment definitions are similar to those reported. We use the current HSR-treatment definition as the main measure for several reasons. First, it gives us one definition across all tests and therefore facilities the interpretation and comparison of results across tests. Second, requiring direct city-to-city HSR connections will leave out many observations, including cities that are connected to HSR. It is also not clear that direct city-to-city connection is a significantly more powerful measure because transferring between HSR lines are usually convenient. Third, if the current HSR-treatment definition is less powerful, it actually biases against our finding significant results.

[21] -0.055/1.235 ≈ -4.5%. -0.005/0.066 ≈ -7.6%.

plotted in Figure 5, shows no HSR effect before connections, but significant effects occur post-connections.

[Table 7 & Figure 5 here.]

4.4. Stock Crash Risk

We next examine whether HSR connections reduce stock crash risk. The results, in Table 8, show a significantly negative treatment effect of HSR connections on crash risk. The coefficient estimates suggest that negative skewness and downside volatility are respectively lower by 7% and 6% of one standard deviation.[22] These results suggest that there is a degree of negative-information concealment before HSR connections, and that HSR connections significantly reduce insiders' abilities to conceal negative information from outside investors. Thus, HSR connections may also have a positive externality on corporate governance. The parallel-trend test, plotted in Figure 6, again shows no HSR effect before connections, but significant effects occur post-connections.

[Table 8 & Figure 6 here.]

4.5. Synchronicity – the Composite Information-Production Measure

As discussed earlier, stock price is the ultimate aggregator of all information, and we use stock price synchronicity as a composite measure of information production. The results are in Table 9. We start in Column 1 without including any control variables. The effect is highly significant and very large: HSR connections reduce synchronicity by about 25% of one standard deviation.[23] But we know there are other determinants of synchronicity, such as firm performance and economic development, so we control for these determinants in Column 2 to address two concerns. The first is that these determinants may have simultaneously caused HSR connections and increased information production. The second concern is that HSR connections may have caused economic growth, which have attracted

---

[22] -0.049/0.691 ≈ -7%. -0.028/0.467 ≈ -6%.

[23] -0.088/0.349 ≈ -25%.

investor attention and led to greater information production. If so, increased information production is not an information effect but rather an economic-growth effect.

The results in Column 2 show that the economic-growth effect certainly exists. For example, firms with higher *ROE*, *Growth*, and *M/B* and in cities with faster GDP growth experience greater information production (lower synchronicity). But the information effect is distinct from the economic-growth effect: the coefficient estimate on *PostHSR* barely changes.

In Column 3, we further control for unobservable time-invariant firm heterogeneity as well as unobservable time-varying shocks. The HSR effect drops in magnitude but remains highly significant: the coefficient estimate implies that HSR connections reduce synchronicity by about 6% of one standard deviation.[24] The regression R^2 increases from 7% to 34%. The above granular inspection of the results in Table 9 suggests that the positive effect of HSR on information production is not spuriously caused by omitted firm performance or economic growth, and that HSR's information effect is distinct from the effects of firm performance or economic growth. We again conduct the parallel-trend test and plot the result in Figure 7. There is no HSR effect before connections, but significant effects occur post-connections.

[Table 9 & Figure 7 here.]

4.6. The Counterfactuals of the HSR Treatment Effect

Even with the DiD framework, to definitively conclude causality is challenging because the HSR buildout is not random, and therefore we cannot observe the counterfactuals. That is, for firms that are connected to HSR, we cannot observe what would have happened if they were not connected to HSR (or if they are disconnected from HSR). A severe HSR accident represents a natural experiment that allows us to at least partly observe the counterfactuals of HSR.

[24] -0.021/0.349 ≈ -6%.

As described in the introduction, on July 23, 2011, lightning struck a section of HSR's power network in Zhejiang Province and caused a massive control system malfunction. As a result, a fast-moving train rear-ended into a slow-moving train, killing 40 passengers. In addition to attributing the accident to lightning strike, ensuing investigations concluded that human errors following the lightning strike also contributed to the disaster. As a result, 54 high-level officials were punished. Zhijun Liu, the head of Ministry of Railways, was investigated for corruption charges and later received a death sentence (with reprieve).

The consequences of this accident were wide-ranging. HSR speed was significantly reduced across the country. The reduced speed diminished the advantages of HSR relative to other forms of transportation. More importantly, the accident shattered the public's confidence in HSR's safety and deterred many people from riding HSR – passenger-per-km, after increasing by 51% from 2009 to 2010, dropped by 4% from 2011 to 2012. This accident provides an opportunity for us to assess the counterfactuals of HSR. That is, what if the already-built HSR are taken down? We answer this question in Table 10.

[Table 10 here.]

In Panel A, we start with institutional-investor activities. Relative to the baseline regression in Table 4, we add a two-way interaction term between *PostHSR* and an indicator variable for 2011 and 2012. The coefficient estimates are large and significantly negative. Comparing the coefficients with those of *PostHSR* suggests that the positive HSR effect on institutional investors' information-production activities is almost completely reversed right after the accident. The three-way interaction term with an indicator for Zhejiang Province (*D_Z*), where the accident occurred, examines whether the reversal effect is more significant for Zhejiang Province. The three-way interaction term is statistically significant for institutional site visits, but not for institutional holdings or the number of institutional investors.

In Panels B to F, we repeat the test in Panel A for the news media, analysts, stock liquidity, stock crash risk, and stock price synchronicity. The findings are consistent across these tests: HSR-induced information production is significantly reversed after the accident.

These counterfactuals of HSR connections further confirm the causal effect of HSR connections on information production.

We also investigate whether other economic shocks in 2011 and 2012 have spuriously caused the reversal effect in 2011 and 2012. We obtain China's GDP data from the World Bank.[25] Between our sample period of 2007 to 2018, China's annual GDP growth rates have been slowly trending downwards, with the average being 8.7%. The annual GDP growth rates for 2011 and 2012 also average to be 8.7%. Thus, it is unlikely that other large negative economic shocks have caused the reversal effect on information production, or that the accident has caused significant economic damage. The accident-produced counterfactuals are additional evidence that the information effect of HSR is independent from the effect of economic growth.

4.7. Is It Really an Information Effect?

Is the HSR effect really an information effect or an effect through other economic channels? In other words, is the HSR effect on information production the result of greater information discovery and dissemination because people can travel faster and more conveniently than before, or is it the result of greater economic growth post HSR attracting more investor attention and hence encouraging greater information production? We first note that these alternative channels are not mutually exclusive and are not orthogonal to each other. What we have shown is that the information effect is robust to controlling for an extensive list of proxies for firm performance and economic development.

Additionally, the passenger-only nature of HSR helps to more sharply identify the information effect since products do not travel on HSR. The counterfactual evidence from the HSR accident is also consistent with the interpretation of an information effect. It is unlikely that the accident significantly impaired the growth prospects of the firms along the HSR lines, and as mentioned earlier, there was no significant damage to GDP. Yet

[25] https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=CN

information production significantly reversed after the accident, consistent with the interpretation of an information effect rather than an economic-growth effect.

If the HSR connections cause an information effect for firms along the HSR lines, we should expect the effect to be driven by firms that have relatively poor ex-ante information environments. This is exactly what we find from the following tests. Our first measure of information environment is the level of unexplained accounting accruals. We run a modified Jones (1991) model and obtain the regression residual. Higher absolute values of the residual plausibly represent a relatively poor information environment for investors, and the HSR's information effect should be more pronounced for these firms. We split the sample based on annual median values of the absolute unexplained accruals. We run the baseline regression of Table 9 and present the result in the first two columns of Table 13. Consistent with our expectation, the HSR effect is significant only for firms with high unexplained accruals.

Second, firms that are audited by large and reputable auditing firms should have better information environments. In contrast, firms with relatively small auditors should have relatively worse information environments, and the HSR's information effect should be more pronounced for these firms. We split the sample based on whether a firm has a big auditor, defined as one of the Big 4 international accounting firms or the Top 10 China-based accounting firms based on the sum of total revenue audited in a given year. Consistent with our expectation, Columns 3 and 4 of Table 13 show that the HSR effect is significant only for firms that use relatively small auditors.

Third, we split the sample based on the level of Internet penetration in a firm's headquarters' city. A lower Internet penetration level suggests relatively poor information infrastructure and worse information environments, and we should detect a more pronounced information effect from the HSR connections. Columns 5 and 6 show exactly that. The HSR effect is driven by firms in low-Internet-penetration cities.

Collectively, the evidence is consistent with the interpretation that HSR connections lead to a significant information effect.

4.8. To Further Address Endogeneity Concerns

The collective evidence so far suggests that HSR connections lead to increased information production. The rest of the paper conducts further tests to address endogeneity concerns, alternative explanations, and the robustness of the findings. Given the large set of such tests and to maintain exposition concision, we will focus on stock price synchronicity, the composite measure of information production. The effects of all information-production channels, including those that are difficult to quantify and not explicitly examined, should be aggregated into stock prices.

4.8.1. Inconsequential-Places Approach

Although the HSR buildout was driven by macro policy makers and therefore largely exogenous to individual cities or firms, the policy makers did not randomly make the HSR plan. Deliberate considerations went into the planning. For example, the 4×4 (and later 8×8) grid is designed to connect large cities, which are usually more politically important and economically advanced. Thus, for these large cities, it is possible that HSR connections and information production are both correlated with economic development, and that all the control variables still cannot eliminate this simultaneity concern.

We first want to note that as mentioned earlier, HSR plans are laid out well in advance of the actual construction. Second, as discussed earlier, non-random assignments of HSR connections do not necessarily mean that the assignments are not exogenous to individual firms. In other words, even though a city is not randomly chosen to be connected to HSR, the connection can still be exogenous to firms in that city because the firms have no influence on whether they are to be connected to HSR. The economics and finance literature have regularly employed non-random but exogenous shocks, such as passages of laws or regulations (e.g., Bertrand and Mullainathan, 2003). Of course, it is possible that because the treatment is not random (even though it is exogenous), the effect of the treatment might be different than the effect of a randomly assigned treatment. In our setting, since HSR is deliberately designed to run through larger cities, how will it bias our findings? Since larger cities already have more developed information environments ex ante, HSR's incremental information effect will be less pronounced than that for smaller cities, whose ex ante information environments are poorer. That is, including larger cities in our sample reduces the test power and biases *against* our finding significant HSR treatment effect, which is exactly what we find from the following test.

We eliminate larger cities from our sample and focus on smaller cities, which are "inconsequential" in the sense that economic development in these smaller cities has little bearing on the decision of HSR connections. In other words, these smaller cities just happen to be along the HSR lines, and thus the treatment effect from HSR is more exogenous and can be more sharply identified. Studies in the infrastructure-economics literature have employed the inconsequential-places approach as an identification strategy to infer causality of infrastructure projects on various economic outcomes (e.g., Chandra and Thompson, 2000; Michaels, 2008).

We run Equation 1 again and progressively eliminate large cities from the sample. The results are in Table 11. In Column 1, we exclude firms in the four municipalities that are under direct administration of the central government, which are Beijing, Shanghai, Tianjin, and Chongqing. All of these are mega-cities with populations over 10 million. Compared to Column 3 of Table 9, the sample size drops by about 5,000, but the coefficient estimate on *PostHSR* grows from -0.021 to -0.024. In Column 2 we further exclude firms in provincial capitals. The sample size drops by another 7,000, but the coefficient estimate on *PostHSR* grows even larger to -0.029. In Column 3, we further eliminate the five cities with "independent planning status".[26] All these cities are economically advanced and are fiscally independent from their provincial governments and report directly to the central government. The sample size drops further by 3,000, but the coefficient estimate on *PostHSR* again grows larger to -0.031. Results from this inconsequential-places approach further support the causal interpretation of HSR connections on information production.

[Table 11 here.]

[26] These five cities are Dalian, Qingdao, Ningbo, Shenzhen, and Xiamen.

4.8.2. Instrumental-Variable (IV) Estimation

The planning and construction of HSR is affected by geographical features. For example, hilly terrains make HSR construction more technically challenging and economically costly. Gradient is a measure in geography that describes the hilliness or slope of a terrain – greater gradient means more hilliness and steeper slopes. Gradient is a truly exogenous feature and has a direct impact on the likelihood of HSR connections, thus satisfying the relevance condition as an IV. Gradient also does not change over time and is therefore unlikely to affect the time-series changes in information production. That is, conditional on the control variables that are present in the regression model, gradient is unlikely to affect information production except through the effect on HSR connections, satisfying the exclusion restriction as an IV. As mentioned in the data section, we collect high-density elevation data, compute the gradient for each city, and use the value of its natural logarithm as an IV for HSR connections.

Panel A of Table 12 reports the IV regression result. As expected, in the first-stage regression of *PostHSR* on gradient, gradient has a significantly negative coefficient estimate. That is, cities with hilly terrains are less likely to be connected to HSR. The Anderson-Rubin test rejects the null hypothesis that gradient is a weak IV. In the second stage, the coefficient estimate on the instrumented *PostHSR* is significantly negative. Thus, the IV regression confirms the baseline DiD regression findings.

[Table 12 here.]

The second instrument we employ for HSR connections is whether a city was connected by a conventional railway in 1962. This "historical route IV" approach is based on the premise that the new rail network can be partly predicted by the historical rail network, yet the historical rail network does not affect the current economic outcome under study except through its effect on the new rail network. Duranton and Turner (2012) develop this IV approach when studying how the U.S. Interstate Highway System affects city-level economic outcomes. They use historical transportation networks as instruments for the interstate highway network at the end of the 20th century. Baum-Snow et al. (2017) study how railway and highway configurations after the 1990s affect economic decentralization in China. They use the urban transportation structure in 1962 as an instrument for the transportation structure after the 1990s.[27] Examples of other studies that employ the "historical route IV" approach include Duranton and Turner (2011), Hsu and Zhang (2014), and Garcia-López, Holl, and Viladecans-Marsal (2015).

Building new HSR lines requires first acquiring and clearing the land, then physically leveling and grading a railbed. It is much more cost-effective to build new lines alongside existing lines. In fact, that is what happened with HSR. Much of the high-speed rail lines are built next to conventional railway lines. Cities that were connected by railways in 1962 likely have geographical features that are more favorable for HSR construction, such as less hilly terrains. Therefore, whether a city was connected to a railway in 1962 is clearly relevant to HSR connections in the 2000s. Furthermore, whether a city was connected to a railway in 1962 is unlikely to affect the temporal changes in local firms' information production in the 2000s. In 1962, China's economy was mostly based in agriculture, of Soviet style, and inwardly focused. Concerns about geopolitical risks also called for a shift of economic resources and development from coastal regions inland. In contrast, China's economy in the 2000s is industrialized, export oriented, and more market driven. Therefore, railway connections in 1962 likely satisfy the exclusion restriction. That is, conditional on the control variables that are present in the regression model, railway connections in 1962 are unlikely to affect information production in the 2000s except through the effect on HSR connections.

The second IV result is in Panel B of Table 12. The first-stage result shows that railway connections in 1962 positively predict HSR connections. The second-stage regression shows that the instrumented *PostHSR* has a significantly negative effect on synchronicity.[28]

---

[27] We follow Baum-Snow et al. (2017) and use 1962 railway information as an instrument. It seems that they choose the year 1962 more for data convenience rather than for any particular event that separates the pre- and post-1962 periods. We do identify one important event in 1961 – in August 1961, China built its first electrified railway (versus the steam or diesel locomotives). The validity of this instrument does not rely on a particular year but rather that it satisfies the relevance condition and the exclusion restriction.

[28] In the second-stage regressions, the coefficient estimates on *PostHSR_HAT* are much larger than those in the baseline regressions in Table 9. The main reason is that the two IVs are noisy instruments in the sense that their partial R^2s in the first-stage regressions are low (Jiang 2017). This is not surprising since the IVs

4.8.3. Falsification Tests

Another endogeneity concern is that correlations in underlying trends may have produced the HSR effect. For our setting, we know that economic development in China, and hence information production, has been trending upwards over time. We also know that over time, more firms have been connected to HSR. Therefore, the concern is that these underlying trends may have only led to the appearance of an HSR effect rather than a true HSR effect. We first note that if the underlying trends affect the treatment and the control groups equally, then the DiD approach already accounts for the effects of underlying trends by design. If the underlying trends affect the treatment and the control groups differently, they may bias the results against us or in our favor. To further address this concern, we conduct two falsification tests. If the HSR effect is truly caused by HSR connections, we should not observe the effect outside the treatment group. To implement this test, we generate two "placebo" treatment variables: placebo cities and placebo firms.

For each year, based on the actual percentage of cities that are connected to HSR (as in Table 1 Panel B), we randomly designate a group of cities as connected to HSR. Similarly, based on the actual percentage of firms that are connected to HSR, we randomly designate a group of firms that are connected to HSR. Once a city or a firm is designated as treated, it remains treated for the subsequent years. We replace *PostHSR* with the placebo variables and re-run the baseline regression in Column 3 of Table 9. Neither of the placebo variables are significant. The results are in Table A1 of the Online Appendix. To assess the robustness of the placebo test, we simulate the placebo regressions 1,000 times. The average coefficient estimates for the placebo variables are essentially zero. Whether we simulate for 500 times, 1500 times, or 2000 times does not change the inference.

---

are static, but *PostHSR* is dynamic. The main takeaway is that the IV regressions confirm the baseline results directionally.

4.8.4. Matched Sample

We also test whether the results are robust to a matched-sample procedure. Each year, we use the full set of control variables to match a treated observation with a control observation based on the propensity score (PS). Alternatively, we create a matched sample using the entropy balancing (EB) procedure (e.g., Hainmueller, 2012; McMullin and Schonberger, 2020). An advantage of EB is that covariate distributions between the treated group and the control group can be better matched through a reweighting process. Table A2 of the Online Appendix reports regression results using matched data sets. The HSR effect persists. Compared with the baseline regression in Table 9, the coefficient estimates on *PostHSR* become larger and remain highly significant.

4.8.5. Controlling for Other Transportation Infrastructures

Another set of potentially omitted correlated variables are other transportation infrastructures, such as highways, waterways, airlines, and conventional railways. It is possible that it is the development of these transportation infrastructures, not HSR, that has led to greater information production, although this alternative explanation does not conflict with our main hypothesis that transportation infrastructure affects information production. When controlling for other transportation infrastructures, the HSR effect remains large and highly significant. The results are in Table A3 of the Online Appendix.

We conduct several additional robustness checks, which we describe in the Online Appendix. Collectively, the empirical results provide consistent evidence that HSR connections lead to greater information production.

5. Conclusion

Back to the opening question of this study: To an investor in New York City who is considering an investment opportunity in Chicago, does it matter whether the train takes 4 hours or 21 hours to get to Chicago? The investor can read everything there is about the investment opportunity, email the management, talk to them over the phone, and meet with them via video conferencing. But by travelling to Chicago and being physically in the context of the information when collecting and processing it, the investor can gain an understanding and perspective that may otherwise be missed. Now imagine thousands of financial-market professionals, along with millions of other people, gaining understandings and perspectives by travelling between places, and sharing them purposely or casually with other millions of people. It is by facilitating this information-production process that transportation infrastructure affects financial markets, which we show in this study.

Empirically, we identify several information channels, such as institutional investors, the media, and stock analysts. We show that HSR's information effect is distinct from the effects of economic growth or firm performance. There are likely other channels at work. For example, as a massive number of people move rapidly on HSR, the word-of-mouth channel may also contribute to HSR's information effect. As more data become available, we will be able to learn about these other channels. The effects of HSR also go beyond information or financial markets. However, the information effect is a first-order effect since information directs the flow of most, if not all, economic resources. For example, information directs the flow of capital; capital formation directs the flow of materials and labor; and the flow of materials and labor leads to economic output. Information, as stated earlier, is the most valuable commodity.

There has been a growing consensus in the U.S. that the country has fallen behind on infrastructure investment. The $1.2 trillion infrastructure bill, signed into law on Nov. 15, 2021, promises to make the largest investment in public transportation infrastructure in recent U.S. history. Hopefully, the findings of this study will contribute to a better understanding of the externalities of transportation infrastructure on financial markets. It is notable that in this era of political polarization, the need to boost infrastructure investment has received broad bipartisan support. Against this backdrop, infrastructure-and-financial-markets seems to be an underexplored but fruitful topic for future research.

References:

Akerlof, G.A., 1970. The market for "lemons": Quality uncertainty and the market mechanism. Quarterly Journal of Economics 84, 488-500.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31-56.

Ansar, A., Flyvbjerg, B., Budzier, A., Lunn, D., 2016. Does infrastructure investment lead to economic growth or economic fragility? Evidence from China. Oxford Review of Economic Policy 32, 360-390.

Asher, S. and Novosad, P., 2020. Rural roads and local economic development. American Economic Review 110, 797-823.

Asquith, P., Mikhail, M.B., Au, A.S., 2005. Information content of equity analyst reports. Journal of Financial Economics 75, 245-282.

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

Baum-Snow, N., 2007. Did highways cause suburbanization?. The Quarterly Journal of Economics 122, 775-805.

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

Bertrand, M., Mullainathan, S., 2003. Enjoying the quiet life? Corporate governance and managerial preferences. Journal of Political Economy 111, 1043-1075.

Chan, K., Chan, Y.C., 2014. Price informativeness and stock return synchronicity: Evidence from the pricing of seasoned equity offerings. Journal of Financial Economics 114, 36-53.

Chandra, A., Thompson, E., 2000. Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. Regional Science and Urban Economics 30, 457-490.

Chen, T., Harford, J., Lin, C., 2015. Do analysts matter for governance? Evidence from natural experiments. Journal of financial Economics 115, 383-410.

Chen, J., Hong, H.G., Stein, J.C., 2001. Forecasting crashes: Trading volume, past returns and conditional skewness in stock prices. Journal of Financial Economics 61, 345-381.

Chen, K, Guo, W., Kang, Y., Xiong, X., Yang, M., 2021. Does time-space compression affect analysts' forecast performance? Working paper, https://ssrn.com/abstract=3839578.

Chen, D., Ma, Y., Martin, X., Michaely, R., 2021. On the fast track: Information acquisition costs and information production. Journal of Financial Economics, forthcoming.

Cheng Q, Du F, Wang X, Wang Y., 2016. Seeing is believing: Analysts' corporate site visits. Review of Accounting Studies 21, 1245-1286.

Coval, J.D., Moskowitz, T.J., 1999. Home bias at home: Local equity preference in domestic portfolios. Journal of Finance 54, 2045-2073.

Coval, J.D., Moskowitz, T.J., 2001. The geography of investment: Informed trading and asset prices. Journal of Political Economy 109, 811-841.

Da, Z., Gurun, U.G., Li, B., Warachka, M., 2021. Investment in a smaller world: The implications of air travel for investors and firms. Management Science 67, 417-435.

Dasgupta, S., Gan, J., Gao, N., 2010. Transparency, price informativeness, and stock return synchronicity: Theory and evidence. Journal of Financial and Quantitative Analysis 45, 1189-1220.

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

Donaldson, D., 2018. Railroads of the Raj: Estimating the impact of transportation infrastructure. American Economic Review 108, 899-934.

Duarte, J., Han, X., Harford, J., Young, L., 2008. Information asymmetry, information dissemination and the effect of regulation FD on the cost of capital. Journal of Financial Economics 87, 24-44.

Duranton, G., Turner, M., 2011. The fundamental law of road congestion: Evidence from U.S. cities. American Economic Review 101, 2616-2652.

Duranton, G., Turner, M., 2012. Urban growth and transportation. Review of Economic Studies 79, 1407-1440.

Durnev, A., Morck, R., Yeung, B., 2004. Value-enhancing capital budgeting and firm-specific stock return variation. Journal of Finance 59, 65-105.

Ellis, J. A., Madureira, L., Underwood, S., 2020. The causal effects of proximity on investment: Evidence from flight introductions. Journal of Financial and Quantitative Analysis 55, 1978-2004.

Engelberg, J.E., Parsons, C.A., 2011. The causal impact of media in financial markets. Journal of Finance 66, 67-97.

Faber, B., 2014. Trade integration, market size, and industrialization: Evidence from China's national trunk highway system. Review of Economic Studies 81, 1046-1070.

Faccio, M., Xu, J., 2015. Taxes and capital structure. Journal of Financial and Quantitative Analysis 50, 277-300.

Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. Journal of Finance 64, 2023-2052.

Gao, M., Huang, J., 2020. Informing the market: The effect of modern information technologies on information production. Review of Financial Studies 33, 1367-1411.

Garcia-López, M.Á., Viladecans-Marsal, E., Holl, A., 2015. Suburbanization and the highways: When the Romans, the Bourbons and the first cars still shape Spanish cities. Journal of Urban Economics 85, 52–67.

Giroud, X., 2013. Proximity and investment: Evidence from plant-level data. Quarterly Journal of Economics 128, 861-915.

Gul, F.A., Kim, J.B., Qiu, A.A., 2010. Ownership concentration, foreign shareholding, audit volume, and stock price synchronicity: Evidence from China. Journal of Financial Economics 95, 425-442.

Hainmueller, J., 2012. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political Analysis 20, 25–46.

Hayek, F.A., 1945. The use of knowledge in society. American Economic Review 35, 519-530.

Hölmstrom, B., 1979. Moral hazard and observability. Bell Journal of Economics, 74-91.

Hsu, W.T., Zhang, H., 2014. The fundamental law of highway congestion: evidence from Japanese expressways. Journal of Urban Economics 81, 65-76.

Hutton, A.P., Marcus, A.J., Tehranian, H., 2009. Opaque financial reports, R^2, and crash risk. Journal of Financial Economics 94, 67-86.

Jiang, W., 2017. Have instrumental variables brought us closer to the truth. Review of Corporate Finance Studies 6(2), 127-140.

Jin, L., Myers, S.C., 2006. R^2 around the world: New theory and new tests. Journal of Financial Economics 79, 257-292.

Jones, J.J., 1991. Earnings management during import relief investigations. Journal of Accounting Research 29, 193–228.

John, K., Knyazeva, A., Knyazeva, D., 2011. Does geography matter? Firm location and corporate payout policy. Journal of Financial Economics 101, 533-551.

Kothari, S.P., Shu, S., Wysocki, P.D., 2009. Do managers withhold bad news? Journal of Accounting Research 47, 241-276.

Kong, D., Liu, L., Liu, S., 2020. Market information traveling on high-speed rails: The case of analyst forecasts. Pacific-Basin Finance Journal, 101320.

Krippner, G.R., 2005. The financialization of the American economy. Socio-Economic Review 3, 173-208.

Kuang, C., Liu, Z., Zhu, W., 2021. Need for speed: High-speed rail and firm performance. Journal of Corporate Finance 66, 101830.

Lawrence, M., Bullock, R., Liu, Z., 2019. China's High-Speed Rail Development. The World Bank, Washington.

Liberti, J.M., Petersen, M.A., 2019. Information: Hard and soft. Review of Corporate Finance Studies 8, 1-41.

Loughran, T., Schultz, P., 2005. Liquidity: Urban versus rural firms. Journal of Financial Economics 78, 341-374.

McMullin, J.L., Schonberger, B., 2020. Entropy-balanced accruals. Review of Accounting Studies 25, 84–119.

Michaels, G., 2008. The effect of trade on the demand for skill: Evidence from the interstate highway system. Review of Economics and Statistics 90, 683-701.

Morck, R., Yeung, B., Yu, W., 2000. The information content of stock markets: Why do emerging markets have synchronous stock price movements? Journal of Financial Economics 58, 215-260.

Petersen, M.A., Rajan, R.G., 2002. Does distance still matter? The information revolution in small business lending. Journal of Finance 57, 2533-2570.

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. Review of Financial Studies 22, 435-480.

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

Shiller, R.J., 2017. Narrative economics. American Economic Review, 107, 967-1004.

Shleifer, A., Vishny, R.W., 1986. Large shareholders and corporate control. Journal of Political Economy 94, 461-488.

Soltes, E., 2014. Private interaction between firm management and sell-side analysts. Journal of Accounting Research 52, 245–272.

Van der Zwan, N., 2014. Making sense of financialization. Socio-Economic Review 12, 99-129.

West, K.D., 1988. Dividend innovations and stock price volatility. Econometrica 56, 37-61.

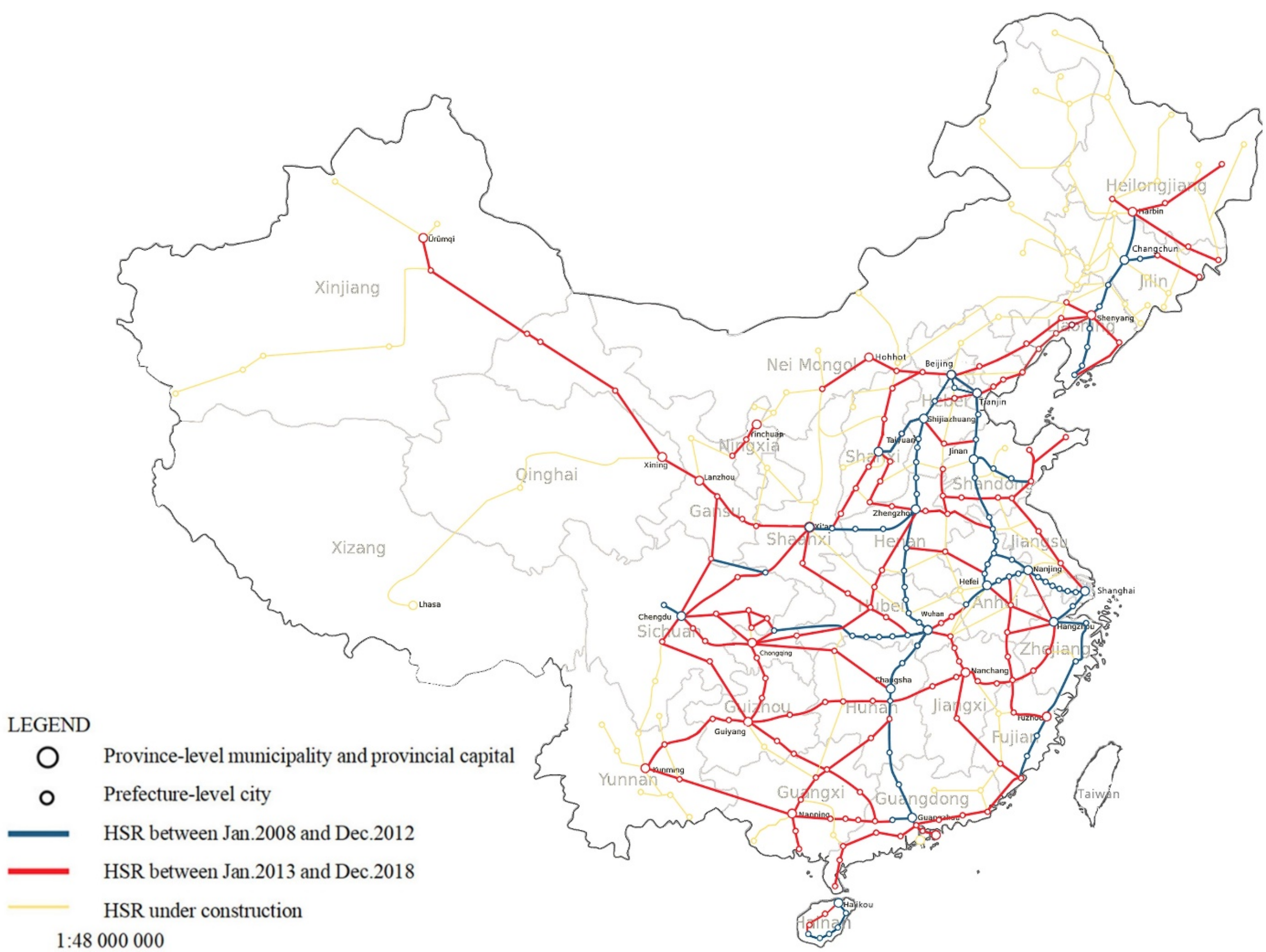
Xiong, J., Ouyang, C., Tong, J.Y., Zhang, F.F., 2021. Fraud commitment in a smaller world: Evidence from a natural experiment. Journal of Corporate Finance 70, 102090.

Appendix A: Variable Definitions

Type	Variable	Definition	Data Source
Dependent variables	*I_ Visit*	The number of site visits by institutional investors	CNRDS(Chinese Research Data Services Platform)
	I_Hold	The percentage of firm equity held by institutional investors	CNRDS
	I_Number	The number of institutional investors	CNRDS
	M_ Visit	The number of site visits by the media	CNRDS
	NetNews	The number of internet-based news that mentions the company	CNRDS
	PaperNews	The number of newspaper-based news that mentions the company	CNRDS
	A_ Visit	The number of site visits by analysts	CNRDS
	A_Follow	The number of analysts that follow the firm	CNRDS
	A_Report	The number of analyst research reports covering the firm	CNRDS
	ILLIQ	We follow Amihud (2002) and measure the daily price impact as the ratio of absolute stock return to trading volume, that is, the price impact per Chinese yuan of trading volume: $Illiquidity_d = \|return_d\|/(price \times volume)_d$. The annual measure is then calculated as: $ILLIQ = \ln(1 + 10^8 \times the\ average\ of\ Illiquidity_d\ in\ a\ given\ year)$.	CSMAR
	Bid-Ask Spread	The average of a firm's daily bid-ask spread. *Bid-Ask Spread*=100×(ask-bid)/[(ask+bid)/2].	CSMAR
	NCSKEW	The negative conditional skewness of weekly stock return Wi,τ in year t, calculated as: $NCSKEW_{i,t} = -[n(n-1)^{\frac{3}{2}} \sum W_{i,\tau}^3] / [(n-1)(n-2)(\sum W_{i,\tau}^2)^{\frac{3}{2}}]$, where subscripts i, τ and t denote firm i, week τ and year t. Wi,τ is the firm-specific weekly return for firm i in week τ in year t, which is measured by the natural log of one plus the residual return based on the market model.	CSMAR (China Stock Market and Accounting Research Database)
	DUVOL	The down-to-up volatility of weekly stock return Wi,τ in the year t, calculated as: $DUVOL_{i,t} = Ln[\sum_{Down} \frac{W_{i,\tau}^2}{(n_d - 1)} / \sum_{Up} \frac{W_{i,\tau}^2}{(n_u - 1)}]$, where n_d and n_u are the number of down weeks and the number of up weeks	CSMAR
	SYNCH	$SYNCH_{i,t} = log[{R_{i,t}}^2/(1 - {R_{i,t}}^2)]$, where R^2 is obtained from annual regressions of stock weekly returns on market returns and industry returns (Equation 2 in the paper).	CSMAR
Treatment variable	*PostHSR*	An indicator variable equal to 1 if a firm's headquarters city is connected to HSR; 0 otherwise. We use web-crawling techniques to collect the dates that each city was connected to an HSR line. We manually clean the data and construct a set of HSR connection dates for each city.	The official website of China Railway State Group.
Control variables	*Size*	The natural logarithm of total assets, adjusted for inflation	CSMAR / NBS
	Age	The natural logarithm of one plus the number of years since the company went public	CSMAR

Type	Variable	Definition	Data Source
	ROE	Net profit divided by equity	CSMAR
	Growth	The growth rate of operating income	CSMAR
	MB	Market-to-book ratio	CSMAR
	R&D	R&D expense, scaled by sales	CSMAR
	R&D_Missing	An indicator variable equal to 1 when R&D value is missing; 0 otherwise.	CSMAR
	Lev	Total debt divided by total assets	CSMAR
	Top1	Percentage stockholding by the top institutional investor	CSMAR
	BigAuditor	An indicator variable equal to 1 if the auditor is a Big 4 international accounting firm or a China-based firm that ranks in the top 10 based on the sum of total revenue audited in year t; 0 otherwise.	CSMAR/CNRDS
	SOE	An indicator variable equal to 1 for state-owned enterprises; 0 otherwise	CSMAR
	TUR	Stock trading volume divided by shares outstanding	CSMAR
	GDP	GDP per capita at the city level, adjusted for inflation	NBS / Wind
	Open	The natural logarithm of the amount of foreign capital utilized by the city, adjusted for inflation	NBS / Wind
	TRA	Road area per capita at the city level	NBS / Wind
	SGDS	The natural logarithm of total retail sales of consumer goods at the city level, adjusted for inflation	NBS / Wind
	POP	The natural logarithm of each city's registered population	NBS / Wind
IVs	*LnGrad*	The natural logarithm of the gradient for each city. We collect 90-meter-resolution elevation data, process it with ArcGIS software to construct an elevation map, and compute each city's gradient.	The Geospatial Data Cloud by China Academy of Sciences.
	Railway1962	An indicator variable equal to 1 if a city was connected to a railway in 1962; 0 otherwise	Baidu Encyclopedia

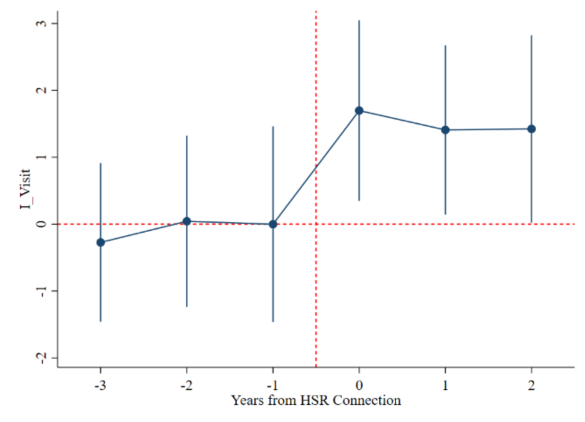
Figure 1: The development of China's HSR network in different stages



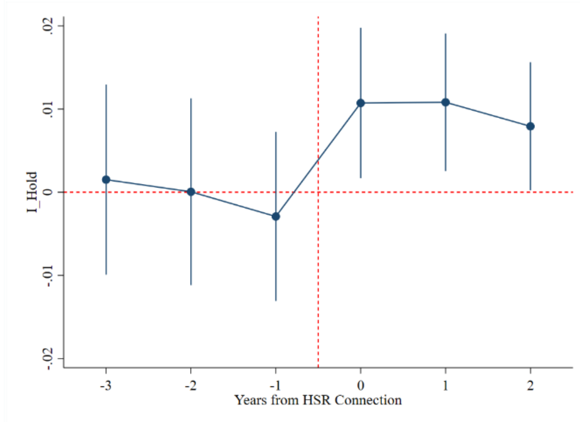
Development Stages	Development Details	Number of cities newly connected to HSR	Operating mileage (km)
Jan. 2008 to Dec. 2012	The 4×4 (four east-west and four north-south) HSR grid was preliminarily completed.	92	8,548
Jan. 2013 to Dec. 2018	The 8×8 HSR grid was under construction.	118	20,548
Jan. 2019 to Dec. 2025	New HSR lines under construction.	61	8,500

Notes: The definition of HSR is: passenger-dedicated railways with a speed of more than 250 km/h and an initial operating speed of more than 200 km/h. The number of cities connected to HSR includes provincial-level municipalities, provincial capitals, and prefecture-level cities. According to the latest mid- to long-term plan for China's railway network, the operating mileage will reach 38,000 km by 2025.

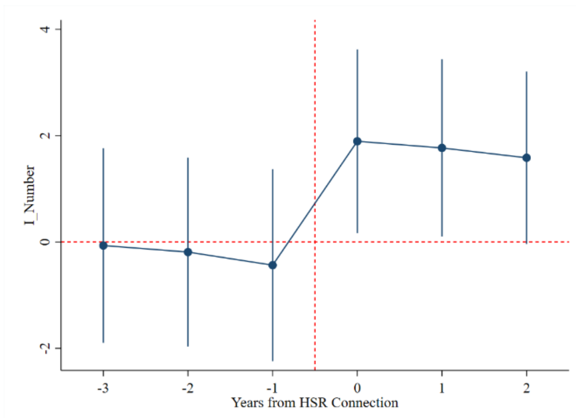
Figure 2: Parallel-Trend Tests for Institutional-Investor Activities



A: Number of Site Visits by Institutional Investors



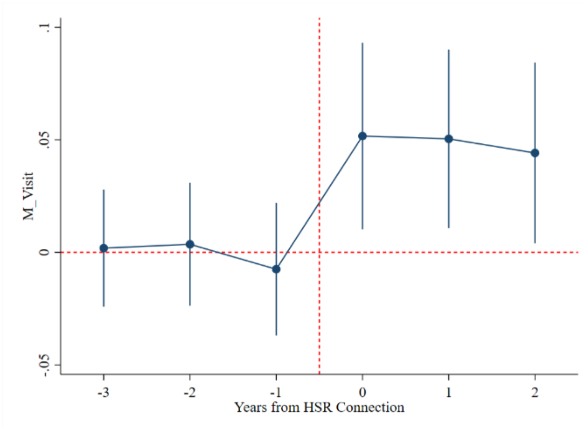
B: Holdings of Institutional Investors



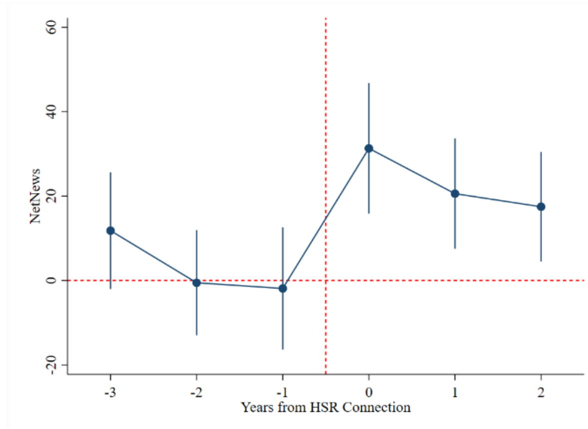
C: Number of Institutional Investors

Note: We implement a regression similar to Equation 1 but replace the PostHSR dummy with six separate dummies for each of the three years before the HSR connection, the connection year, and each of the two years after the connection. Plotted are the coefficient estimates for the six dummies. The dependent variables are institutional-investor activities. All control variables in the baseline regression are included. All specifications are OLS regressions with firm and year fixed effects. None of the three coefficient estimates pre-HSR is statistically significant. The coefficient estimates for the connection year and the two post-HSR years are statistically significant at 10% level or better.

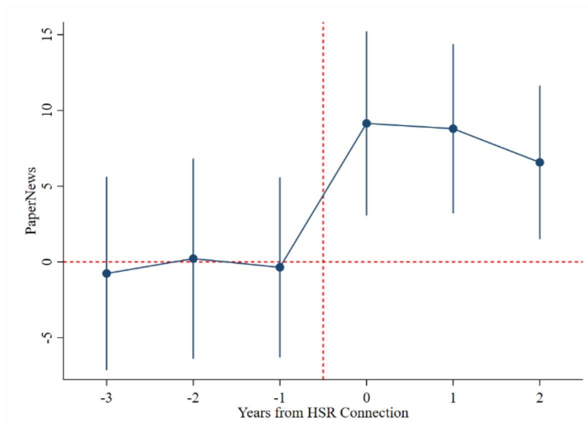
Figure 3: Parallel-Trend Tests for Media Activities



A: Number of Site Visits by the Media



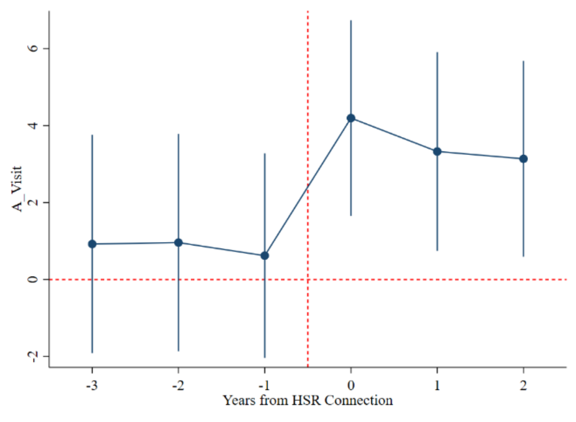
B: Number of Internet News Stories



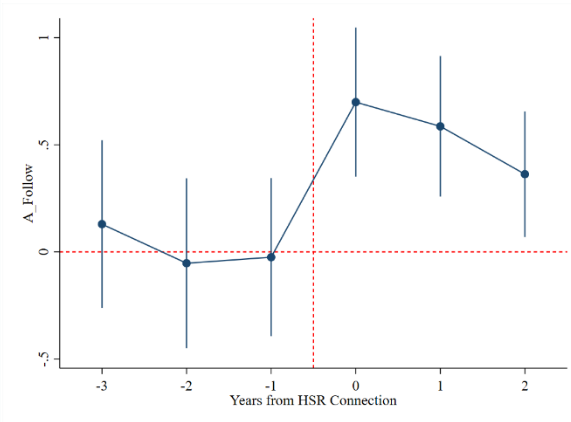
C: Number of Newspaper Stories

Note: We implement a regression similar to Equation 1 but replace the PostHSR dummy with six separate dummies for each of the three years before the HSR connection, the connection year, and each of the two years after the connection. Plotted are the coefficient estimates for the six dummies. The dependent variables are media activities. All control variables in the baseline regression are included. All specifications are OLS regressions with firm and year fixed effects. None of the three coefficient estimates pre-HSR is statistically significant. The coefficient estimates for the connection year and the two post-HSR years are statistically significant at 10% level or better.

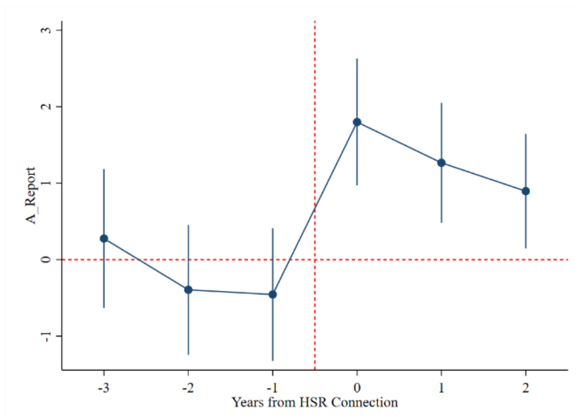
Figure 4: Parallel-Trend Tests for Analyst Activities



A: Number of Site Visits by Analysts



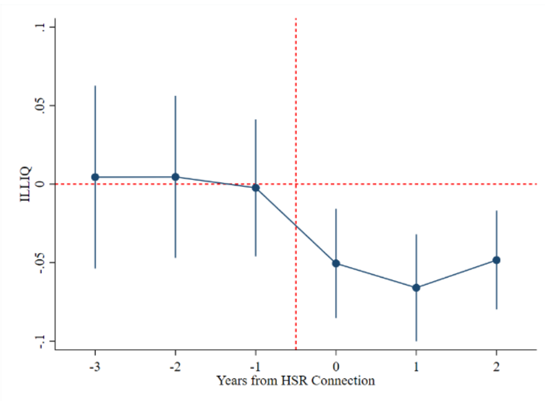
B: Number of Analyst Following



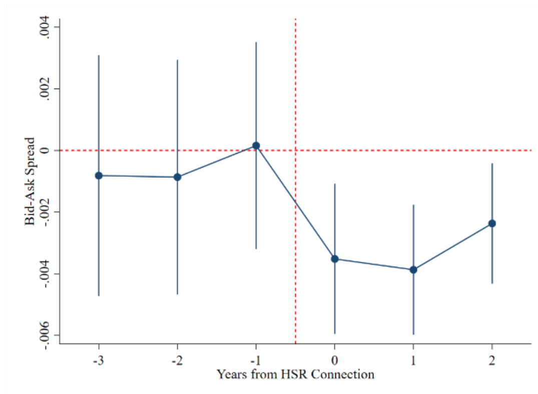
C: Number of Analyst Reports

Note: We implement a regression similar to Equation 1 but replace the PostHSR dummy with six separate dummies for each of the three years before the HSR connection, the connection year, and each of the two years after the connection. Plotted are the coefficient estimates for the six dummies. The dependent variables are analyst activities. All control variables in the baseline regression are included. All specifications are OLS regressions with firm and year fixed effects. None of the three coefficient estimates pre-HSR is statistically significant. The coefficient estimates for the connection year and the two post-HSR years are statistically significant at 10% level or better.

Figure 5: Parallel-Trend Tests for Stock Liquidity



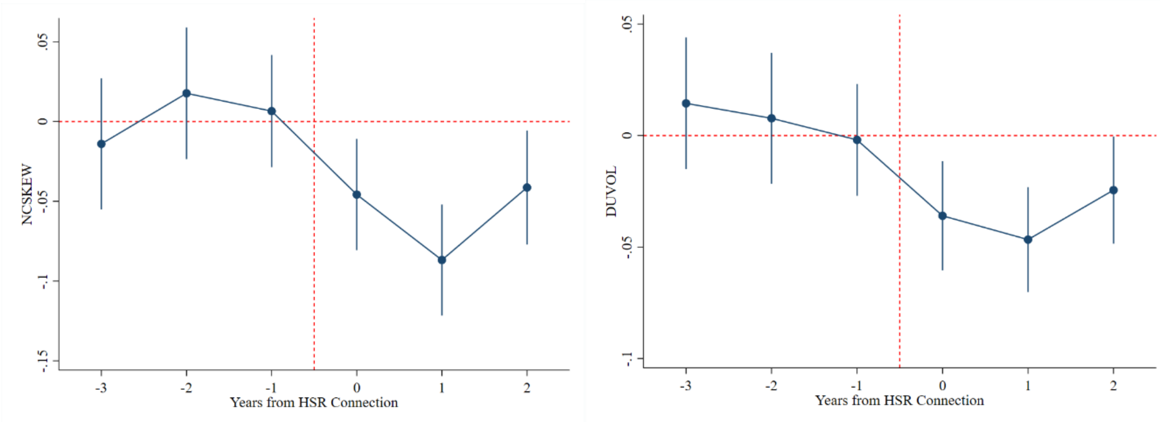
A: ILLIQ



B: Bid-Ask Spread

Note: We implement a regression similar to Equation 1 but replace the PostHSR dummy with six separate dummies for each of the three years before the HSR connection, the connection year, and each of the two years after the connection. Plotted are the coefficient estimates for the six dummies. The dependent variables are ILLIQ and Bid-Ask Spread. All control variables in the baseline regression are included. All specifications are OLS regressions with firm and year fixed effects. None of the three coefficient estimates pre-HSR is statistically significant. The coefficient estimates for the connection year and the two post-HSR years are statistically significant at 10% level or better.

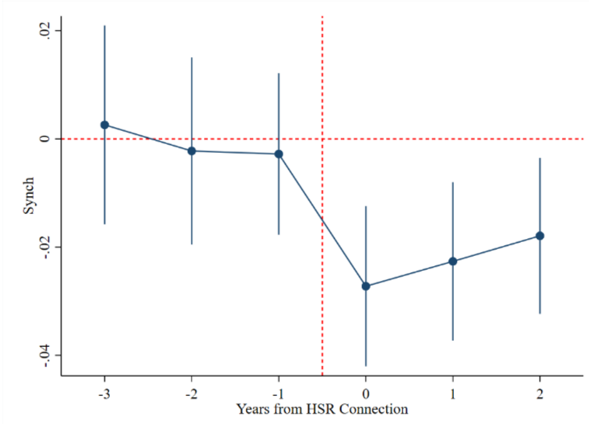
Figure 6: Parallel-Trend Tests for Stock Crash Risk



A: Negative Skewness B: Down-to-Up Volatility

Note: We implement a regression similar to Equation 1 but replace the PostHSR dummy with six separate dummies for each of the three years before the HSR connection, the connection year, and each of the two years after the connection. Plotted are the coefficient estimates for the six dummies. The dependent variables are measures for stock crash risk. All control variables in the baseline regression are included. All specifications are OLS regressions with firm and year fixed effects. None of the three coefficient estimates pre-HSR is statistically significant. The coefficient estimates for the connection year and the two post-HSR years are statistically significant at 10% level or better.

Figure 7: Parallel-Trend Test for Stock Price Synchronicity



Note: We implement a regression similar to Equation 1 but replace the PostHSR dummy with six separate dummies for each of the three years before the HSR connection, the connection year, and each of the two years after the connection. Plotted are the coefficient estimates for the six dummies. The dependent variable is stock price synchronicity. All control variables in the baseline regression are included. All specifications are OLS regressions with firm and year fixed effects. None of the three coefficient estimates pre-HSR is statistically significant. The coefficient estimates for the connection year and the two post-HSR years are statistically significant at 10% level or better.

Table 1: Summary Statistics and Annual Distribution

This table presents summary statistics for the main variables in the baseline regression. The sample comprises 25,600 firm-year observations of 2,999 firms in 240 cities between 2007 and 2018. All variables are defined in Appendix A. *I_Visit*, *M_Visit*, and *A_Visit* are only available for firms listed on the Shenzhen Stock Exchange. In Panel B, Column 1 shows the annual number of observations that are treated by HSR connections (*PostHSR*=1). Column 2 shows the total number of observations, and Column 3 is the percentage of observations that has treated. Similarly, Columns 4 to 6 shows the treatment statistics at the city level.

Panel A: Summary Statistics						
Variables	N	Mean	Std.	Min	Median	Max
I_Visit	15268	14.875	26.961	0.000	0.000	96.000
I_Hold	25600	0.366	0.238	0.000	0.363	0.874
I_Number	25600	37.947	49.814	1.000	18.000	273.000
M_Visit	15268	0.119	0.564	0.000	0.000	4.000
NetNews	25600	265.918	377.968	6.000	150.000	2572.000
PaperNews	25600	73.460	163.467	1.000	25.000	1169.000
A_Visit	15268	28.102	47.858	0.000	8.000	276.000
A_Follow	25600	7.376	9.090	0.000	4.000	40.000
A_Report	25600	14.777	20.947	0.000	6.000	101.000
ILLIQ	25600	-3.092	1.235	-5.775	-3.167	1.628
Bid-Ask Spread	25600	0.160	0.066	0.060	0.147	0.422
NCSKEW	25600	-0.301	0.691	-2.381	-0.266	1.640
DUVOL	25600	-0.203	0.467	-1.365	-0.201	1.016
R^2	25600	0.476	0.175	0.085	0.483	0.822
SYNCH	25600	-0.053	0.349	-1.033	-0.030	0.665
PostHSR	25600	0.726	0.446	0.000	1.000	1.000
Size	25600	21.823	1.286	19.254	21.658	25.766
Age	25600	2.053	0.866	0.000	2.197	3.219
ROE	25600	0.065	0.126	-0.691	0.071	0.378
Growth	25600	0.213	0.529	-0.580	0.118	3.738
MB	25600	2.111	1.781	0.201	1.646	10.532
R&D	25600	0.030	0.042	0.000	0.018	0.251
R&D_Missing	25600	0.263	0.441	0.000	0.000	1.000
Lev	25600	0.438	0.215	0.048	0.433	0.960
Top1	25600	0.354	0.150	0.087	0.335	0.750
BigAuditor	25600	0.226	0.418	0.000	0.000	1.000
SOE	25600	0.420	0.494	0.000	0.000	1.000
TUR	25600	6.524	4.940	0.587	5.114	25.122
GDP	25600	6.764	3.063	1.287	6.728	14.372
Open	25600	41.193	51.067	0.089	23.965	300.000
TRA	25600	14.974	8.623	4.080	12.720	44.570
SGDS	25600	7.480	1.087	4.605	7.650	9.105
POP	25600	6.393	0.666	4.580	6.469	8.115

Panel B: Annual Distribution						
Year	No. of treated observations (1)	Total No. of observations (2)	% of treated observations (3) = (1) / (2)	No. of treated cities (4)	Total No. of cities (5)	% of treated cities (6) = (4) / (5)
2007	0	1,293	0	0	205	0
2008	226	1,376	16.42	15	209	7.17
2009	411	1,431	28.72	33	211	15.64
2010	919	1,744	52.69	52	222	23.42
2011	1,368	2,039	67.09	67	227	29.52
2012	1,604	2,206	72.71	87	230	37.83
2013	1,746	2,216	78.79	102	230	44.35
2014	2,017	2,308	87.39	133	232	57.33
2015	2,244	2,476	90.63	151	234	64.53
2016	2,472	2,656	93.07	164	239	68.62
2017	2,764	2,921	94.63	170	240	70.83
2018	2,802	2,934	95.50	185	240	77.08
Total	18,573	25,600	--	--	--	--

Table 2: Correlation

This table presents correlation coefficients between main variables, which are defined in Appendix A. Below the diagonal are Pearson's correlation coefficients. Above the diagonal are Spearman's rank correlation coefficients. a, b, and c indicate statistical significance at the 1%, 5%, and 10% level.

	Post HSR	*I_ Visit*	*I_ Hold*	*I_ Number*	*M_ Visit*	*Net News*	*Paper News*	*A_ Visit*	*A_ Follow*	*A_ Report*	*ILLIQ*	*Bid-Ask Spread*	*NC SKEW*	*DU VOL*	*SYNCH*
PostHSR	1	0.344^{a}	0.129^{a}	0.238^{a}	0.110^{a}	0.276^{a}	0.014^{b}	0.206^{a}	0.093^{a}	0.110^{a}	-0.187^{a}	-0.190^{a}	-0.025^{a}	-0.027^{a}	-0.096^{a}
I_Visit	0.266^{a}	1	0.099^{a}	0.375^{a}	0.284^{a}	0.304^{a}	0.069^{a}	0.690^{a}	0.316^{a}	0.329^{a}	-0.292^{a}	-0.267^{a}	0.003	-0.009	-0.100^{a}
I_Hold	0.129^{a}	0.112^{a}	1	0.511^{a}	0.010	0.272^{a}	0.254^{a}	0.146^{a}	0.280^{a}	0.280^{a}	-0.247^{a}	-0.075^{a}	0.009	0.001	-0.071^{a}
I_Number	0.184^{a}	0.358^{a}	0.414^{a}	1	0.087^{a}	0.520^{a}	0.396^{a}	0.430^{a}	0.619^{a}	0.615^{a}	-0.625^{a}	-0.397^{a}	0.047^{a}	0.028^{a}	-0.051^{a}
M_Visit	0.104^{a}	0.258^{a}	0.010	0.087^{a}	1	0.127^{a}	0.065^{a}	0.237^{a}	0.079^{a}	0.083^{a}	-0.076^{a}	-0.074^{a}	-0.008	-0.013	-0.060^{a}
NetNews	0.173^{a}	0.211^{a}	0.244^{a}	0.485^{a}	0.109^{a}	1	0.645^{a}	0.327^{a}	0.396^{a}	0.401^{a}	-0.436^{a}	-0.321^{a}	-0.042^{a}	-0.045^{a}	-0.153^{a}
PaperNews	0.077^{a}	0.092^{a}	0.198^{a}	0.365^{a}	0.053^{a}	0.714^{a}	1	0.202^{a}	0.353^{a}	0.348^{a}	-0.320^{a}	-0.177^{a}	-0.025^{a}	-0.024^{a}	-0.079^{a}
A_Visit	0.193^{a}	0.731^{a}	0.148^{a}	0.464^{a}	0.216^{a}	0.252^{a}	0.161^{a}	1	0.471^{a}	0.469^{a}	-0.285^{a}	-0.272^{a}	0.018^{b}	0.001	-0.056^{a}
A_Follow	0.079^{a}	0.330^{a}	0.285^{a}	0.595^{a}	0.064^{a}	0.359^{a}	0.289^{a}	0.468^{a}	1	0.938^{a}	-0.367^{a}	-0.359^{a}	0.072^{a}	0.059^{a}	-0.057^{a}
A_Report	0.114^{a}	0.340^{a}	0.273^{a}	0.604^{a}	0.071^{a}	0.386^{a}	0.302^{a}	0.461^{a}	0.872^{a}	1	-0.367^{a}	-0.349^{a}	0.066^{a}	0.054^{a}	-0.066^{a}
ILLIQ	-0.179^{a}	-0.264^{a}	-0.253^{a}	-0.543^{a}	-0.076^{a}	-0.364^{a}	-0.250^{a}	-0.259^{a}	-0.329^{a}	-0.326^{a}	1	0.574^{a}	0.006	0.021^{a}	0.068^{a}
Bid-Ask Spread	-0.199^{a}	-0.241^{a}	-0.067^{a}	-0.291^{a}	-0.062^{a}	-0.170^{a}	-0.092^{a}	-0.219^{a}	-0.275^{a}	-0.254^{a}	0.443^{a}	1	0.004	0.010	0.111^{a}
NCSKEW	-0.035^{a}	0.016^{c}	0.007	0.053^{a}	-0.009	-0.020^{a}	-0.012^{c}	0.037^{a}	0.095^{a}	0.087^{a}	-0.027^{a}	0.006	1	0.842^{a}	0.108^{a}
DUVOL	-0.036^{a}	0.000	-0.000	0.041^{a}	-0.010	-0.027^{a}	-0.014^{b}	0.020^{b}	0.088^{a}	0.080^{a}	-0.002	0.016^{b}	0.796^{a}	1	0.092^{a}
SYNCH	-0.104^{a}	-0.099^{a}	-0.074^{a}	-0.117^{a}	-0.055^{a}	-0.118^{a}	-0.061^{a}	-0.089^{a}	-0.084^{a}	-0.106^{a}	0.035^{a}	0.116^{a}	0.112^{a}	0.089^{a}	1

Table 3: Univariate Comparisons

This table compares the mean and the median of variables between the treatment group and the control group. All variables are defined in Appendix A. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

	PostHSR =1 (Treatment Group)		*PostHSR* =0 (Control Group)		*Diff in Mean*	*Diff in Median*
	Mean	Median	Mean	Median		
I_Visit	18.986	2.000	2.515	0.000	16.471***	2.000***
I_Hold	0.385	0.392	0.317	0.291	0.069***	0.101***
I_Number	43.504	21.000	23.259	10.000	20.245***	11.000***
M_Visit	0.153	0.000	0.018	0.000	0.135***	0.000***
NetNews	306.639	176.000	158.288	84.000	148.352***	92.000***
PaperNews	81.226	25.000	52.932	25.000	28.294***	-0.000
A_Visit	33.442	12.000	12.049	2.000	21.393***	10.000***
A_Follow	7.767	4.000	6.342	3.000	1.425***	1.000***
A_Report	16.174	7.000	11.087	4.000	5.086***	3.000***
ILLIQ	-3.216	-3.288	-2.764	-2.849	-0.451***	-0.439***
Bid-Ask Spread	0.152	0.140	0.182	0.165	-0.030***	-0.025***
NCSKEW	-0.316	-0.274	-0.262	-0.248	-0.054***	-0.026***
DUVOL	-0.213	-0.207	-0.175	-0.191	-0.037***	-0.016**
R^2	0.464	0.471	0.507	0.514	-0.043***	-0.043***
SYNCH	-0.077	-0.051	0.011	0.025	-0.088***	-0.076***
			Control Variables:			
Size	21.903	21.733	21.609	21.477	0.294***	0.256***
Age	2.051	2.197	2.059	2.398	-0.008	-0.201***
ROE	0.065	0.071	0.066	0.072	-0.001	-0.001
Growth	0.209	0.116	0.223	0.124	-0.013*	-0.008**
MB	2.116	1.650	2.097	1.628	0.019	0.022
R&D	0.035	0.028	0.013	0.001	0.021***	0.027***
R&D_Missing	0.196	0.000	0.442	0.000	-0.246***	-0.000***
Lev	0.426	0.419	0.470	0.470	-0.044***	-0.051***
Top1	0.352	0.333	0.357	0.342	-0.005**	-0.009***
BigAuditor	0.236	0.000	0.200	0.000	0.025***	0.000***
SOE	0.382	0.000	0.520	1.000	-0.139***	-1.000***
TUR	6.179	4.649	7.437	6.447	-1.258***	-1.802***
GDP	7.673	7.550	4.361	3.826	3.313***	3.728***
Open	50.321	34.132	17.065	6.160	33.256***	27.972***
TRA	15.565	13.220	13.412	11.010	2.153***	2.190***
SGDS	7.845	7.972	6.516	6.477	1.329***	1.480***
POP	6.500	6.541	6.111	6.142	0.389***	0.398***

Table 4: HSR and Information Production by Institutional Investors

This table presents regression results of Equation 1. The dependent variables are the number of institutional investor site visits, percent of institutional investor holdings, and the number of institutional investors. *PostHSR* is an indicator variable equal to 1 if a firm's headquarters city is connected to HSR, and 0 otherwise. All specifications are OLS regressions with firm and year fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: HSR connections increase institutional investor activities.

Dep. Var. =	*I_Visit*	*I_Hold*	*I_Number*
	(1)	(2)	(3)
PostHSR	1.913**	0.012**	1.557**
	(2.366)	(2.136)	(2.147)
Size	6.747***	0.036***	20.097***
	(9.892)	(8.843)	(29.169)
Age	4.048***	0.002	-2.238***
	(5.241)	(0.350)	(-2.645)
ROE	12.001***	0.059***	21.493***
	(6.688)	(6.052)	(12.886)
Growth	0.753*	-0.001	-2.298***
	(1.931)	(-0.541)	(-7.393)
MB	0.967***	0.022***	4.369***
	(4.746)	(16.939)	(21.849)
R&D	26.146**	0.189***	36.568***
	(2.384)	(2.873)	(3.848)
R&D_Missing	1.117	-0.010*	-1.818**
	(1.310)	(-1.819)	(-2.354)
Lev	-4.437*	-0.007	-4.410*
	(-1.846)	(-0.481)	(-1.947)
Top1	-9.737**	0.112***	-20.709***
	(-2.171)	(4.592)	(-5.004)
BigAuditor	-0.457	-0.001	1.295***
	(-0.841)	(-0.210)	(2.696)
SOE	1.849	0.016	0.986
	(0.974)	(1.506)	(0.575)
TUR	0.414***	-0.013***	-0.409***
	(7.175)	(-31.111)	(-7.273)
GDP	1.251***	-0.005**	0.899***
	(3.425)	(-2.204)	(2.834)
Open	0.011	0.000	0.027***
	(0.848)	(1.413)	(3.590)
TRA	0.090	-0.001	-0.239**
	(0.650)	(-1.400)	(-2.093)
SGDS	-3.373	-0.003	-4.591

Dep. Var. =	*I_Visit*	*I_Hold*	*I_Number*
	(1)	(2)	(3)
	(-0.896)	(-0.142)	(-1.509)
POP	5.509	-0.051**	-2.679
	(1.256)	(-2.002)	(-0.648)
Year FE	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*
N	15268	25600	25600
Adj-R^2	0.422	0.608	0.692

Table 5: HSR and Information Production by the Media

This table presents regression results of Equation 1. The dependent variables are the number of site visits by the media, the number of Internet news stories reported by the media, and the number of newspaper stories published by the media. *PostHSR* is an indicator variable equal to 1 if a firm's headquarters city is connected to HSR, and 0 otherwise. All specifications are OLS regressions with firm and year fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: HSR connections increase media attention.

Dep. Var. =	*M_Visit*	*NetNews*	*PaperNews*
	(1)	(2)	(3)
PostHSR	0.038**	23.202***	8.513***
	(2.384)	(2.626)	(2.658)
Controls	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*
N	15268	25600	25600
Adj-R^2	0.102	0.665	0.717

Table 6: HSR and Information Production by Analysts

This table presents regression results of Equation 1. The dependent variables are the number of analyst site visits, the number of analyst following, and the number of analyst reports. *PostHSR* is an indicator variable equal to 1 if a firm's headquarters city is connected to HSR, and 0 otherwise. All specifications are OLS regressions with firm and year fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: HSR connections increase analyst activities.

Dep. Var. =	*A_Visit*	*A_Follow*	*A_Report*
	(1)	(2)	(3)
PostHSR	3.822***	0.635***	2.201***
	(2.730)	(3.001)	(4.603)
Controls	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*
N	15268	25600	25600
Adj-R^2	0.431	0.616	0.544

Table 7: HSR and Stock Liquidity

This table presents regression results of Equation 1. The dependent variable is the Amihud (2002) illiquidity measure and the bid-ask spread. The independent variable of interest is *PostHSR*. All specifications are OLS regressions with year and firm fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: HSR connections enhances stock liquidity.

Dep. Var. =	*ILLIQ*	*Bid-Ask Spread*
	(1)	(2)
PostHSR	-0.055***	-0.005***
	(-2.618)	(-3.301)
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
N	25600	25600
Adj-R^2	0.584	0.550

Table 8: HSR and Stock Crash Risk

This table presents regression results of Equation 1. The dependent variable is the negative conditional skewness of weekly stock returns (NCSKEW) or the down-to-up volatility of weekly stock returns (DUVOL). The independent variable of interest is *PostHSR*. All specifications are OLS regressions with year and firm fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: HSR connections reduce stock crash risk.

Dep. Var. =	*NCSKEW*	*DUVOL*
	(1)	(2)
PostHSR	-0.049***	-0.028**
	(-2.805)	(-2.354)
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
N	24330	24330
Adj-R^2	0.080	0.073

Table 9: Synchronicity – the Composite Information-Production Measure

This table presents regression results of Equation 1. The dependent variable is stock price synchronicity. *PostHSR* is an indicator variable equal to 1 if a firm's headquarters city is connected to HSR, and 0 otherwise. All specifications are OLS regressions. Year and firm fixed effects are included where indicated. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: HSR connections reduce stock price synchronicity, and the effect is independent from the effects of firm performance and local economic development.

Dep. Var. =	*SYNCH*	*SYNCH*	*SYNCH*
	(1)	(2)	(3)
PostHSR	-0.088***	-0.087***	-0.021***
	(-15.418)	(-13.254)	(-2.788)
Size		-0.010***	-0.010
		(-3.374)	(-1.635)
Age		-0.005	0.009
		(-1.285)	(1.105)
ROE		-0.099***	-0.075***
		(-4.832)	(-3.831)
Growth		-0.046***	-0.034***
		(-10.977)	(-8.817)
MB		-0.041***	-0.045***
		(-21.063)	(-19.750)
R&D		0.566***	0.282***
		(8.249)	(3.060)
R&D_Missing		0.060***	0.028***
		(9.366)	(3.535)
Lev		-0.167***	-0.173***
		(-11.629)	(-8.284)
Top1		-0.043**	0.105***
		(-2.305)	(2.851)
BigAuditor		-0.005	0.012**
		(-0.893)	(2.113)
SOE		0.060***	0.021
		(9.460)	(1.194)
TUR		-0.001*	-0.010***
		(-1.701)	(-15.466)
GDP		-0.004**	0.005
		(-2.150)	(1.570)
Open		0.000**	0.000**
		(1.967)	(2.235)
TRA		0.000	0.000
		(0.785)	(0.407)
SGDS		0.009	0.033
		(1.134)	(1.235)

Dep. Var. =	*SYNCH*	*SYNCH*	*SYNCH*
	(1)	(2)	(3)
POP		-0.000	-0.005
		(-0.002)	(-0.147)
Year FE	*NO*	*NO*	*YES*
Firm FE	*NO*	*NO*	*YES*
N	25600	25600	25600
Adj-R^2	0.013	0.071	0.338

Table 10: The Counterfactuals: Evidence from an HSR Accident

The dependent variables are the information-production measures examined in previous tables. The independent variables of interest are the two-way interaction term between *PostHSR* and the 2011-to-2012 indicator (*D_1112*) as well as the three-way interaction term that includes an indicator for Zhejiang Province (*D_Z*), where in July 2011, a fatal accident killed 40 passengers, causing nationwide speed reductions and damaged public confidence in HSR's safety. The interaction term between *PostHSR* and *D_1112* captures the counterfactuals of HSR connections. The three-way interaction term with *D_Z* examines whether the accident's effects are nationwide or are localized to Zhejiang Province. All control variables in the baseline regression are included. All specifications are OLS regressions with firm and year fixed effects except for Panel D Column 2, which is a logit regression and the marginal effect is reported. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: The HSR's treatment effect is significantly reversed after the accident, as indicated by the two-way interaction term's coefficients that are of the opposite sign to the coefficients of *PostHSR*. Coefficient estimates of the three-way interaction term are mostly insignificant, meaning that the accident's effects are nationwide and not localized to Zhejiang Province.

Panel A: Institutional-investor activities

Dep. Var. =	*I_Visit*	*I_Hold*	*I_Number*
	(1)	(2)	(3)
PostHSR	3.059***	0.017***	2.261**
	(3.367)	(2.815)	(1.988)
PostHSR × *D_1112*	-4.434***	-0.019***	-2.837**
	(-5.234)	(-2.654)	(-2.006)
PostHSR × *D_1112* × *D_Z*	-6.159**	0.040	-1.276
	(-1.997)	(1.539)	(-0.324)
These terms are included but not reported: *D_1112, D_Z, PostHSR* × *D_Z,* and *D_1112* × *D_Z.*			
Controls	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*
N	15268	25600	25600
Adj-R²	0.423	0.618	0.648

Panel B: Media activities

Dep. Var. =	*M_ Visit*	*NetNews*	*PaperNews*
	(1)	(2)	(3)
PostHSR	0.059***	33.291***	12.532***
	(3.397)	(3.286)	(3.283)
PostHSR × D_1112	-0.064***	-26.797***	-12.488***
	(-3.519)	(-2.694)	(-2.915)
PostHSR × D_1112 × D_Z	-0.089	46.653	14.853
	(-1.438)	(1.567)	(1.004)
These terms are included but not reported: *D_1112, D_Z, PostHSR × D_Z,* and *D_1112 × D_Z.*			
Controls	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*
N	15268	25600	25600
Adj-R^2	0.102	0.666	0.717

Panel C: Analyst activities

Dep. Var. =	*A_ Visit*	*A_Follow*	*A_Report*
	(1)	(2)	(3)
PostHSR	4.451***	0.764***	2.758***
	(2.607)	(3.356)	(5.261)
PostHSR × D_1112	-9.924***	-0.687***	-1.848***
	(-3.848)	(-2.694)	(-2.830)
PostHSR × D_1112 × D_Z	-6.400	-0.605	-2.192
	(-1.079)	(-0.706)	(-1.087)
These terms are included but not reported: *D_1112, D_Z, PostHSR × D_Z,* and *D_1112 × D_Z.*			
Controls	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*
N	15268	25600	25600
Adj-R^2	0.440	0.616	0.544

Panel D: Stock liquidity

Dep. Var. =	*ILLIQ*	*Bid-Ask Spread*
	(1)	(2)
PostHSR	-0.098***	-0.009***
	(-3.968)	(-4.568)
PostHSR × *D_1112*	0.143***	0.013***
	(4.407)	(5.451)
PostHSR × *D_1112* × *D_Z*	0.023	0.007
	(0.222)	(0.965)
These terms are included but not reported: *D_1112, D_Z, PostHSR* × *D_Z,* and *D_1112* × *D_Z.*		
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
N	25600	25600
Adj-R^2	0.584	0.551

Panel E: Stock crash risk

Dep. Var. =	*NCSKEW*	*DUVOL*
	(1)	(2)
PostHSR	-0.071***	-0.051***
	(-3.583)	(-3.819)
PostHSR × *D_1112*	0.066**	0.054***
	(2.226)	(2.619)
PostHSR × *D_1112* × *D_Z*	-0.021	-0.061
	(-0.242)	(-0.928)
These terms are included but not reported: *D_1112, D_Z, PostHSR* × *D_Z,* and *D_1112* × *D_Z.*		
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
N	25600	25600
Adj-R^2	0.080	0.074

Panel F: Stock price synchronicity

Dep. Var. =	*SYNCH*
PostHSR	-0.025***
	(-2.985)
PostHSR × *D_1112*	0.035***
	(2.956)
PostHSR × *D_1112* × *D_Z*	-0.068*
	(-1.804)
These terms are included but not reported: *D_1112, D_Z, PostHSR* × *D_Z,* and *D_1112* × *D_Z.*	
Controls	*YES*
Year FE	*YES*
Firm FE	*YES*
N	25600
Adj-R^2	0.338

Table 11: Excluding Major Cities

We re-run the baseline regression and progressively exclude large or economically importantly cities from the sample. The dependent variable is stock price synchronicity. The independent variable of interest is *PostHSR*. All control variables in the baseline regression are included. All specifications are OLS regressions with year and firm fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: The HSR treatment effect is not concentrated in large or economically important cities. This mitigates the endogeneity concern that economic development in large cities spuriously caused the correlation between HSR connections and information production.

Dep. Var. = *SYNCH*	Excluding the four municipalities that are under direct administration of the central government	Further excluding all provincial capitals	Further excluding the five cities with "independent planning status"
	(1)	(2)	(3)
PostHSR	-0.024***	-0.029***	-0.031***
	(-2.891)	(-2.765)	(-2.585)
Controls	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*
N	20405	13368	10170
Adj-R^2	0.334	0.335	0.340

Table 12: Instrumental Variable Regressions

We employ two instrumental variables for *PostHSR*. The first one is the natural logarithm of a city's gradient (*LnGrad*). Gradient is a measure in geography that describes the hilliness or slopes of a terrain – greater gradient indicates more hilliness and steeper slopes. The second instrument is whether a city was connected by a conventional railway in 1962 (*Railway1962*). We employ each instrument in Panels A and B. All specifications are OLS regressions with year and firm fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: Both instruments are strong predictors for *PostHSR* in the first-stage regressions. The instrumented *PostHSR* is significantly negative in the second-stage regressions, consistent with the baseline results.

Panel A: Using *LnGrad* as an instrument for *PostHSR*

	First stage	Second Stage
Dep. Var. =	*PostHSR*	*SYNCH*
	(1)	(2)
LnGrad	-0.042***	
	(-5.030)	
PostHSR _HAT		-1.522***
		(-3.101)
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
Anderson-Rubin Wald test F-stat	15.67	
(p-value)	(0.000)	
N	25600	25600
R^2	0.529	-

Panel B: Using *Railway1962* as an instrument for *PostHSR*

	First stage	Second Stage
Dep. Var. =	*PostHSR*	*SYNCH*
	(1)	(2)
Railway1962	0.104***	
	(3.620)	
PostHSR _HAT		-4.866***
		(-3.593)
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
Anderson-Rubin Wald test F-stat	279.84	
(p-value)	(0.000)	
N	25600	25600
R^2	0.528	-

Table 13: Heterogeneity Tests

This table presents regression results of Equation 1. The dependent variable is stock price synchronicity. *PostHSR* is an indicator variable equal to 1 if a firm's headquarters city is connected to HSR, and 0 otherwise. We run subsample regressions based on a firm's information environment. Firms with lower levels of unexplained accounting accruals, using big auditors, or in cities with high Internet penetration are expected to have better information environment. We run the following modified Jones Model:

$TA_t = \alpha_1 \left(\frac{1}{A_{t-1}}\right) + \alpha_2(\Delta REV_t - \Delta REC_t) + \alpha_3(PPE_t) + \varepsilon_t$, where TA_t is total accruals, ΔREV_t is the change in revenue from year *t-1* to year *t*, ΔREC_t is the change in net receivables from year *t-1* to year *t, and* PPE_t is the gross property, plant, and equipment in year *t*, all scaled by total assets *A* in year *t-1*. The absolute value of the regression residual is denoted as Abs(DACC) and measures the level of unexplained accruals. We split the sample based on the median value of Abs(DACC) each year. The auditor is a BigAuditor if it is a Big 4 international accounting firm or a China-based accounting firm that ranks in the top 10 based on the sum of total revenue audited in a given year. Internet Penetration is measured as the percentage of a city's Internet users in a given year. We split the sample based on the median value of each year. All specifications are OLS regressions. Year and firm fixed effects are included. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: The HSR treatment effect is driven by firms with relatively poor information environment, suggesting that HSR connections lead to a significant information effect.

Dep. Var. = *SYNCH*	Abs(DACC)		BigAuditor		Internet Penetration	
	Low	High	Yes	No	High	Low
	(1)	(2)	(1)	(2)	(3)	(4)
PostHSR	-0.013	-0.030***	-0.014	-0.028***	0.009	-0.054***
	(-1.158)	(-2.671)	(-0.695)	(-3.180)	(0.831)	(-4.268)
Controls	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
N	12810	12790	5945	19655	12388	13212
R^2	0.351	0.319	0.373	0.335	0.356	0.331

Online Appendix (not to be included in the main paper)

Table A1: Falsification Tests

The dependent variable is stock price synchronicity. The independent variables of interest are the two placebo variables that replace *PostHSR*. First, for each year, based on the actual percentage of cities that are connected to HSR (as in Table 1 Panel B), we randomly designate a group of cities as connected to HSR, and create a dummy variable Placebo_City. Similarly, based on the actual percentage of firms that are connected to HSR, we randomly designate a group of firms that are connected to HSR, and create a dummy variable Placebo_Firm. Once a city or firm is assigned as treated, it remains treated for subsequent years. All control variables in the baseline regression are included. All specifications are OLS regressions with year and firm fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: The HSR treatment effect is not a result of chance or correlations in underlying economic trends.

Dep. Var. = *SYNCH*	(1)	(2)
Placebo_City	-0.030	
	(-1.286)	
Placebo_Firm		-0.001
		(-0.222)
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
N	25600	25600
Adj-R^2	0.337	0.337

Table A2: Matched Samples

The dependent variable is stock price synchronicity. The independent variable of interest is *PostHSR.* Each year, we use the full set of control variables and match a treated observation with a control observation base on propensity-score matching (PSM) or Entropy Balancing (EB) matching. We then run Equation 1 using the matched samples. All control variables in the baseline regression are included. All specifications are OLS regressions with year and firm fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: The HSR treatment effect persists in matched samples.

Dep. Var. = *SYNCH*	PSM	EB
	(1)	(2)
PostHSR	-0.035***	-0.138***
	(-3.294)	(-7.410)
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
N	20592	25600
R^2	0.326	0.301

Table A3: Controlling for Other Transportation Infrastructures

The dependent variable is stock price synchronicity. The independent variable of interest is *PostHSR.* In Panel A we control for the development of other transportation infrastructures, namely passenger volumes of railways, highways, and waterways at the provincial level, whether a city has an airport, and the number of weekly departure flights from a city's airport. In Panel B, we further isolate the HSR effect by excluding cities that have an airport. We do this in two ways. In Column 1, we exclude cities that have HSR connections *and* an airport. That is, we exclude cities that have an airport from the treatment group. In Column 2, we exclude cities that have an airport from our sample. That is, we remove these cities from the treatment group as well as the control group. All control variables in the baseline regression are included. All specifications are OLS regressions with year and firm fixed effects. In parentheses are *t*-statistics that are adjusted for heteroskedasticity and firm-level clustering. ***, **, and * indicate significance at the 1%, 5%, and 10% levels. An intercept is included but not reported. All variable definitions are in Appendix A.

Interpretation: The HSR treatment effect is not some other transportation effect in disguise.

Panel A: Controlling for the development of other transportation infrastructures

Dep. Var. = *SYNCH*	(1)	(2)	(3)	(4)	(5)
PostHSR	-0.021***	-0.021***	-0.021***	-0.021***	-0.022***
	(-2.796)	(-2.799)	(-2.794)	(-2.794)	(-2.907)
Railway	0.005				
	(0.310)				
Highway		0.003			
		(0.328)			
Waterway			-0.001		
			(-0.189)		
Airport				-0.006	
				(-0.095)	
AirPerWeek					-0.025
					(-1.330)
Controls	*YES*	*YES*	*YES*	*YES*	*YES*
Year FE	*YES*	*YES*	*YES*	*YES*	*YES*
Firm FE	*YES*	*YES*	*YES*	*YES*	*YES*
N	25600	25600	25600	25600	25600
Adj-R^2	0.337	0.337	0.337	0.337	0.337

Panel B: Excluding cities that have an airport

Dep. Var. = *SYNCH*	Excluding cities that have HSR connections and an airport (1)	Excluding cities that have an airport (2)
PostHSR	-0.036**	-0.034**
	(-2.262)	(-2.402)
Controls	*YES*	*YES*
Year FE	*YES*	*YES*
Firm FE	*YES*	*YES*
N	10096	5410
Adj-R^2	0.337	0.371

Additional Robustness Checks

In untabulated tests, we perform additional robustness checks. First, as discussed in the data section, as time passes, greater proportions of the data set are treated with HSR connections. As a result, there is less dispersion in the *PostHSR* variable in later years. This reduced dispersion actually reduces the test power and biases against our finding a significant treatment effect. When we limit the sample to 2007-2014 and re-run the baseline regressions, we find similar results. Second, we exclude industry returns from Equation 2 and use only market returns to estimate R^2 and synchronicity. We use this alternative synchronicity measure to re-run the baseline regressions and obtain consistent results. Third, we take the first-difference of all variables and re-run the baseline regression. First-differencing essentially eliminates all time-invariant firm heterogeneity and estimates the HSR treatment effect only for the treatment year. This first-difference regression produces consistent results.

Fourth, there is a possibility that using all firm-year observations may bias the results in our favor. This is because conceivably, high-profile cities are connected to HSR early in the sample period, and firms in these cities will therefore represent a larger proportion of the treatment group. These firms may also have greater information production than firms in low-profile cities, which are connected to HSR late in the sample period. Thus, the observed HSR effect on information production might reflect the fundamental differences between these two sets of firms. Recall that in an earlier test, we excluded large cities and obtained consistent results. To further address this concern, we keep a firm in the sample only between two years before and two years after HSR connection. This keeps each firm's presence in the treatment group and control group balanced. Results from this restricted sample are similar to the baseline results.