On the Fast Track: Information Acquisition Costs and Information Production

Deqiu Chen, Yujing Ma, Xiumin Martin, and Roni Michaely

Abstract
Using the introduction of high-speed rail (HSR) as an exogenous shock to costs of information acquisition, we show that reductions in information-acquisition costs lead to (i) a significant increase in information production, evidenced by a higher frequency of analysts visiting portfolio firms, and (ii) improvement in output quality, manifested in higher forecast accuracy and better recommendations. The effect is more pronounced for firms with information that is difficult to produce. Importantly, more information production is also associated with improved price efficiency. We corroborate these findings using a large-scale survey of financial analysts. Finally, both the empirical and survey results highlight the importance of soft information in analysts' unique-information production.

Keywords: Information acquisition, Acquisition cost, Information quality, Soft information, Sell side analysts, Price efficiency

JEL Codes: D8, G14, G2, G24, G3

Chen is from the University of International Business and Economics, email: nkchendq2006@uibe.edu.cn; Ma is from School of Accounting and Finance, Hong Kong Polytechnic University, email: yu-jing.ma@polyu.edu.hk; Martin is from the Olin Business School, Washington University in St. Louis, email: xmartin@wustl.ed; Michaely is from the Geneva Finance Research Institute, University of Geneva and SFI, email: roni.michaely@unige.ch. This paper has benefited from comments of seminar participants at The Hebrew University of Jerusalem, Ohio State University, Washington University in St. Louis, McMaster University Accounting Research Symposium 2019, Sun Yat-Sen University Finance International Conference 2019, University of Alberta, University of International Business and Economics Conference 2019, Gennaro Bernile, Rich Frankel, Xavier Giroud, Todd Gormley, Ohad Kadan, Mark Leary, Asaf Manela, Phil Dybvig, and Jean-Charles Roche.

Introduction

Using exogenous shocks to the costs of information acquisition and several unique sources of data, we examine three important questions. First, we ask whether information costs affect the amount of information agents collect. Many influential theories (e.g., Grossman and Stiglitz 1980; Verrecchia 1982) predict that the costs of information acquisition affect agents' collection of information. However, due to limited data on *actual* information collection, no direct evidence shows lower acquisition costs result in more information collection. Given the centrality of this assumption in economics and finance (Jensen 2007; Aker 2010; Giroud 2013; Bernstein et al. 2016), illuminating this issue can contribute to the assessment of the validity of these theories. Second, we ask whether lower information-production costs result in not only more information gathering but also higher-quality information. We test this prediction using a precise measure of agents' information quality. Third, we ask whether reductions in information-acquisition costs and thus information asymmetry affect price efficiency.

We employ a unique setting where information-acquisition costs exogenously changed: the introduction of high-speed railways (HSR), which drastically eased travel between cities in China, and consequently allowed analysts easier access to information, especially soft information. This setting allows us to use a novel dataset to gauge the amount of information agents acquired, namely, the number of analysts' site visits to their portfolio companies and, at the same time, to measure the quality of those analysts' information: their earnings forecast accuracy. Based on these measures, we assess the impact of HSR introduction on analysts' information collection and quality. We find HSR introduction results in more information production: Analysts visit their portfolio companies more often. Output quality also improves: Analysts' earnings forecasts are more accurate, and forecast revisions and recommendations convey more information to the market, resulting in greater price efficiency.

Our main identification strategy is a difference-in-differences estimation around the exogenous change in information-acquisition costs, which was staggered from 2008 to 2018, and which can be illustrated with the following example. Meida Nylon, a firm located in Jiangmen, was followed by five analysts from five different brokerages in 2005. The five brokerages are located in Guangzhou, Changsha, Guiyang, Shanghai and Shenzhen. In 2012, Guangzhou-Jiangmen-Zhuhai HSR was introduced, the first HSR connecting the firm to the Guangzhou brokerage, while no HSR connected the firm to the other four brokerages during our sample period. Thus, post 2012, we code the analyst from Guangzhou as a "treatment," and the other four analysts as "controls." We compare the changes in information collection and its quality for the same firm between analysts who have easier access to the firm after HSR introduction and those who did not experience a change in information-collection costs, essentially eliminating many endogeneity concerns.

In spirit, our setting resembles that of Giroud (2013) and Bernstein et al. (2016) with five important distinctions. First, our unique data allow us to explicitly measure agents' information acquisition via site visits to portfolio firms, whereas headquarters managers' plant visits, studied by Giroud, or venture capitalists' visits to portfolio firms, examined by Bernstein et al., are unobserved by researchers. Second, lower travel costs can increase information production by headquarters managers or venture capitalists, and reduce their agency costs with plants or portfolio firms. Giroud's and Bernstein et al.'s settings do not enable them to separate the two effects. By contrast, our setting, using analysts' information production (site visits) and output (forecast errors), allows us to isolate the information effects. Third, our data allow us to provide evidence on another important dimension of reduction in information-acquisition costs—its impact on price efficiency. Fourth, shocks to information-acquisition costs employed in these prior studies are introductions of direct airline routes, similar to our setting of HSR introduction. These events might be endogenous to firms' (Bernstein et al.) and plants' (Giroud) time-varying shocks. Our data allow us to explicitly control for firm- and broker-specific shocks by including firm and broker by year fixed effects. Because of zero variation in the dependent variable within a firm-year (Bernstein et al.) or plant-year (Giroud), Giroud and Bernstein et al. cannot use this structure. Rather, they use a smaller subsample of airline company mergers that is less prone to endogeneity. The fixed-effects structure enables us to use a broad sample, thereby increasing external validity. Finally, our unique setting allows us to assess the type of information being gathered during site visits.

To further investigate the direct impact of the HSR introduction on the amount of information collection and its impact on the overall information quality, we also conduct a large-scale survey of 334

Chinese sell-side analysts. As we explain in more detail below, the survey strengthens our ability to make causal inferences on the relationship between HSR introduction and analyst information production. Importantly, the survey results offer additional insights into the type of information analysts collect during their visits, which is difficult to assess by using archival data alone.

We find treated analysts significantly increased the number of site visits, by 4.9% annually, following HSR introduction. These results cannot be explained by expected higher growth of the firm's city, nor the centrality of the analyst's city, because we include both firm and broker city-time fixed effects in our regressions. The results cannot be attributable to firm- and broker-specific shocks, because our alternative estimation structure also controls for firm and broker by year fixed effects. They also cannot be attributed to lobbying efforts for HSR introduction connecting the firm-broker-pair cities. We also examine whether HSR introduction affects analysts' information production at the extensive margin. We show that analysts increase the likelihood of initiating coverage of a firm to which they are connected post HSR introduction. Our evidence suggests that information cost reduction improves analysts' information production at both intensive margin and extensive margin. Our survey evidence echoes these findings: 96% of respondents agreed they would visit a portfolio company more frequently after the introduction of an HSR route connecting them to the firm. The impact is particularly acute for cities that are harder to reach and for analysts who are more time-constrained. Our results also suggest face-to-face interaction still represents an important source of information, and reductions in information-acquisition costs (travel time) significantly increase the amount of information agents collect.

Second, we test whether HSR introduction affects analysts' information quality, measured by their earnings forecast accuracy. The HSR introduction connecting a firm-analyst pair significantly increases forecast accuracy by 2.1%. Our survey evidence corroborates these findings: 82% of survey respondents agreed or strongly agreed the HSR introduction has helped them make better earnings forecasts. Only 8% thought it would not be likely to have an impact. In sum, the combined evidence suggests reductions in information-acquisition costs result in more information gathering and higher information quality. We further stress-test our results by examining how the HSR effect varies with the difficulty of visiting the portfolio firm without HSR (effectively conducting a triple-difference analysis). We find more pronounced increases in both analyst site visits and forecast accuracy post HSR in cases where analysts have more difficulty visiting firms, suggesting the effect of acquisition costs on information quality likely operates through information production. We also examine the cross-sectional variation in the HSR effect on analyst forecast accuracy along corporate governance and information opacity, both of which affect the difficulty for analysts to access firm information, particularly soft in nature. The evidence suggests the HSR introduction improves forecast accuracy more for firms that are weakly governed and informationally opaque. Additionally, we examine whether site visits and earnings forecast errors trend differently between treated and control analysts before HSR introduction. We don't find evidence of this pattern. Therefore, pre-existing trends cannot explain our results.

Third, we study whether acquisition costs affect price efficiency. Because reductions in information-acquisition costs, due to the HSR introduction, lead to more frequent information collection and better information at the hands of analysts, we expect the market reaction to analyst forecast revisions to be stronger, analyst stock recommendations to be more profitable, and analysts' information to be reflected in prices faster, increasing price efficiency. Indeed, we find a significant increase in investors' reaction to forecast revisions, around forecast revisions by 1.7%, and in the market reaction to stock recommendations by 1.9%, post HSR introduction (measured by three-day abnormal returns). In addition, if prices indeed incorporate information faster following HSR introduction, the information content of earnings announcements will be reduced, because, at this point, more of the information is impounded into prices (see Merkley et al. 2017 for similar arguments). We find the market reaction to a firm's earnings announcements declines significantly following the increase in the intensity of the firm's brokers who are connected to the firm by HSR. Taken in tandem, our evidence of stronger market reaction to analyst research and weaker market reaction to earnings announcements post HSR introduction, suggests reductions in information acquisition costs result in prices impounding information faster thereby improving price efficiency.

Fourth, the evidence that information costs reduction improves output quality implies that this effect operates through information collection. We therefore explicitly estimate the effect of information collection (site visits) on output quality using two-stage least square (2SLS), in which we instrument site visits by HSR introduction. We find site visits enable analysts to increase the quality of information they collect and result in greater price efficiency. From economic perspective, a 1% increase in site visits is associated with a 0.359% reduction in forecast errors, 0.456% increase in the market reaction to forecast revisions, and 0.409% increase in the recommendation profitability. In summary, our evidence suggests that by collecting more information site visits benefit analysts in improving their output quality.

Fifth, we study the type of information analysts gather during their site visits. Our findings suggest soft information is the main culprit. Soft information is subjective and contextual, often depending on face-to-face interaction, and thus its collection should be more sensitive to HSR introduction (Liberti and Petersen 2019). We employ two methods to examine this question. We first test whether the HSR effect varies with the importance of soft information in forecasting performance. Using a measure of the importance of soft information (explained in section 5.1), we find HSR indeed has a stronger effect on both analyst site visits and analyst forecast accuracy among firms for which soft information matters more. Second, we use the survey to examine the role of soft information during site visits. Roughly 87% of survey respondents agree the introduction of HSR helps them better understand items that can be categorized as soft information: corporate culture, employee morale, and firm strategies. These results corroborate the notion that analysts acquire a significant amount of soft information through face-to-face interaction with firms, when soft information plays an important role in assessing future performance. They also suggest that even in an era when technology has significantly affected information collection, human interaction is still valuable.

1. Institutional background of the HSR network

To tackle issues associated with China's fast economic growth, such as the growth of mega cities, congestion, and pollution, the Ministry of Railway (MOR) announced its high-speed rail plan in

November 2006.[1] In 2008, the State Council set the goal of a national HSR grid, with the objective of connecting provincial capitals and other major cities with faster means of transportation. The past 10 years have brought significant growth of China's HSR network, which can be divided into four stages, as seen in Figure 1. In Stage 1 (before August 2008), HSR was not in operation. In Stage 2 (2008/8–2011/7), HSR construction and operation was preliminary, and HSR lines largely concentrated along the east corridor, with total operating mileage of 6,963 km by the end of 2011. In Stage 3 (2011/7−2013/1), the number of HSR lines steadily grew, gradually forming a network with extension to the middle region, reaching a total operating mileage of 12,344 km by the end of 2013. In Stage 4 (2013/1–2018/12), the number of HSR lines grew rapidly, further connecting the northwest region, and the operating mileage at the end of 2018 was 29,000 km, more than double that of Stage 3.[2]

The introduction of HSR drastically eased travel across cities for at least three reasons. First, it reduced travel time substantially. For example, the travel time between Beijing and Shanghai was shortened from 13 hours by traditional train to five by HSR. Second, compared to a flight, HSR reduces travel uncertainty, such as weather or operational delays, and improves connections to inner-city transportation systems. Third, HSR runs more frequently between cities than flights and thus provides travelers with greater flexibility (confirmed in the survey we conducted). An additional benefit is that passengers can use the internet and mobile phones on HSR, which they generally cannot on flights. Research finds the HSR introduction increased labor mobility, particularly for industries with a higher reliance on nonroutine cognitive skills, such as finance (Lin 2017). Because both time and flexibility are important considerations when analysts travel to gather information from firms, HSR introductions are expected to reduce their costs of information collection.

2. Data and variable construction

[1] High-speed trains in China have two main categories. One set of trains starts their numbers with G, running at a designed speed of 350 km/h, and the other set starts their numbers with D, running at a designed speed of 250 km/h. We consider both sets to be HSR in this study, following Lin (2017).

[2] The MOR's criteria for which cities to select to connect to HSR have not be publicized. Therefore, the HSR introduction may be endogenous. Fortunately, our research design uniquely defines treatment by pair locations, by including firm city- time, analyst city-time, and firm-analyst-pair fixed effects. Using an alternative estimation structure of including firm-time and analyst-time fixed effects, and focusing on the introduction of HSR lines that do not involve firm-broker-pair cities located on the two ends of the HSR line, we can effectively alleviate the endogeneity concern. Details on our identification are discussed in section 3.

2.1 Sample

The China Stock Market & Accounting Research (CSMAR) database covers Chinese publicly traded firms listed on Shanghai (SHSE) and Shenzhen (SZSE) stock exchanges. It contains data on analyst annual forecasts and analyst stock recommendations issued by Chinese brokerage houses, stock price, and firm financial information.[3] Because the coverage of analyst forecasts in the CSMAR database is poor before 2005, our sample starts in 2005 and ends in 2019 which provides one-year post-HSR data to examine the HSR effect for rails introduced in 2018. We impose several filters on the data. We exclude firms in the financial industry, due to some brokerages' affiliation with financial firms (e.g., CITIC Bank and CITIC Securities Co. are subsidiaries of CITIC Group) and because this affiliation might reduce information asymmetry and thus the power of the test. We further exclude firms traded on the Second Board Market (Growth Enterprise Market or GEM) and focus on firms on the Main Board, because the GEM firms face different rules and regulations, which might affect investor information demand and thus analysts' research activity differently.[4] We eliminate special treatment (ST) firms, firms in the year of IPO, and a broker-firm pair in which the broker covers the firm for less than a year. We do not require all analysts to be present throughout the sample period for reasons discussed in Internet Appendix IA.1. Since our experiment centers on the effect of the reduction of informational costs, our main sample excludes firms located in mega cities (Beijing, Shanghai, Guangzhou, and Shenzhen), because airline travel is easy for analysts covering these firms, and broker-firms pairs that are in the same city (local analysts). We instead use this portion of the sample to conduct a placebo test. Finally, we restrict the analysis to current-year earnings forecasts, because they are most frequently issued, compared with long-term forecasts. The final analyst-forecasts sample consists of 161,858 earnings forecasts with non-missing EPS data and other data needed to construct control

[3] Chinese firms have announced their earnings quarterly and annually since 2002. Before 2002, Chinese firms announced only semi-annual and annual earnings. Chinese analysts, however, only provide annual earnings forecasts.

[4] For example, the China Securities Regulatory Commission (CSRC) requires investors to have at least two years of trading experience before they can trade GEM stocks, whereas no such requirement exists for the Main Board stocks. In addition, the listing requirements of GEM firms are relatively lax, compared with the requirements of the Main Board firms, which might lead to a differential information environment between the two groups.

variables. These forecasts correspond to 11,381 firm-years with 1,596 distinct firms located in 243 cities and 82 brokerages located in 33 cities, as well as 76,161 and 85,624 forecasts issued to SHSE and SZSE firms, respectively. The larger number of forecasts for the SZSE firms is consistent with the fact that SZSE is a larger exchange (2,100 vs. 1,400 firms), with its listed firms having a higher trading volume. In terms of geographical distribution, brokerages are concentrated in Beijing, Shenzhen, and Shanghai, representing 49% of the total number of brokers and 64% of total earnings forecasts in our sample. Nevertheless, roughly 36% of brokerage activities and all portfolio firms are located outside these areas, allowing us to identify our treatment effect (HSR introduction) defined at the firm-broker-pair-city level.

For a portion of the analyst-forecasts sample, we can obtain brokers' site visits. Through discussions with analysts it became clear that a site visit is commonly conducted at a firm's headquarters commonly involving analysts' face-to-face interactions with divisional managers, rank-and-file employees, investor relations officers, and a tour of operating and production facilities. Therefore, we identify the destination of a site visit by the firm's headquarters. Because analysts typically belong to the research department of a broker commonly located at the broker's headquarters, we identify analysts' location by the headquarters. Starting from 2009, SZSE has required firms to disclose corporate site visits in their annual reports.[5] Then, starting in 2012, SZSE further required firms to provide real-time disclosure of site visits via the official CSRC designated website (www.cninfo.com.cn). We retrieve site visits data from Chinese Research Data Services (CNRDS), which has been used by Agarwal et al. (2020) and Huang et al. (2020). We then merged site-visits data with the analyst-forecasts sample, requiring that a broker issue at least one forecast without a requirement of nonzero site visits for each firm-year during 2009–2019, to form the analyst-site-visits sample. This sample consists of 33,200 firm-broker-year observations for 2009–2019, corresponding to 858 distinct firms located in 191 cities and 80 brokerages located in 33 cities.[6]

[5] "Guidelines on Information Disclosure of Corporate Financial Reports, Document No.1." Available at https://www.szse.cn).
[6] China adopted Fair Disclosure rules, similar to the US, in 2007 (see Article 41 of CSRC passed on December 13, 2006). CSRC further clarifies that private information obtained from site visits, if incorporated in research reports, is illegal. Furthermore, SZSE requires that firms sign memorandum and agreement of site visits certifying

2.2 Variable construction

The exogenous shock to information-acquisition costs that we consider is HSR introduction connecting a firm-city to its broker-city. Specifically, if the HSR route was introduced connecting the firm-city to the broker-city during year *t-1*, *HSR* takes the value of 1 for t and onward, and 0 otherwise. We manually collected HSR route data from China Railway Yearbooks for the period 2003–2019. The data include, for example, the origin and destination stations, train stops, the starting date of operation, and rail type. The detailed data on train stops for each HSR route at each point in time allow us to determine whether and when a firm-broker pair is HSR connected (*HSR*=1). For our analyst-forecasts sample, by the beginning of 2019, 1,454 unique firms (91%), 7,786 firm-years (68%), 175 cities (72%), and 80 brokerages (98%) were connected to the HSR network; 1,159 unique firms (73%) and 6,026 firm-years (53%) were connected to at least one broker by an HSR, and 11,828 (59%) out of 20,068 unique firm-broker pairs were HSR- connected. We observe similar statistics for the analyst-site-visits sample. The significant number of firm-broker pairs that are not HSR connected reinforce the advantage of defining the treatment effect at the firm-broker-pair-city level. As Table 1 shows, 40.5% of firm-broker-years are classified as HSR connected (*HSR*=1) for our site-visits sample (Panel A) and 39.0% for the analyst-forecasts sample (Panel B).

The unique data on actual site visits allow us to estimate changes in analysts' information-acquisition activities. Table 1 Panel A indicates the average broker visits a firm 0.59 times per year.[7] Per our conversations with three major Chinese brokerages, analysts commonly work in teams and the leading analyst signs the research report. Analyst teams in our sample cover 37 firms per year, which means 22 visits occur per team-year (=0.59*37), a nontrivial amount of time required for analysts.[8] The

that firms do not disclose material private information. We do not assume managers provide private information during analyst site visits. Instead, our evidence suggests site visits allow analysts to collect soft information, such as corporate culture, strategy, and employees' morale, which is evident in our survey responses.

[7] The average site visits of 0.59 is lower than it would be if our sample includes firms located in four mega cities and local analysts for which travel costs are lower. For example, we find the average annual frequency of visits is one per firm-broker pair when we focus on local analysts who cover firms located in the four mega cities.

[8] Because a broker commonly has one analyst team to provide coverage for any given firm and our empirical tests are conducted at the broker-firm pair level, we use broker and analyst team (analyst) interchangeably throughout the paper.

number of site visits received by a firm is spread across a fiscal year by visual inspection (untabulated), with no particular concentration in any single month. Firms on average receive coverage by 15 brokers and thus eight broker visits per year (=0.59*15). The second key variable is analyst forecast quality, proxied by *Analyst Forecast Error*, defined as the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year, expressed as a percentage. As shown in Table 1 Panel B, this variable exhibits right skewness with a mean and median of 1.22% and 0.6%, respectively.

We control for a host of firm characteristics in both analyst site visits and forecast-errors analyses. Control variables include firm size, leverage, ROA, book-to-market, return volatility, the number of analysts following, and broker firm-specific experience (consistent with prior studies such as Malloy 2005). Variable definitions are provided in Appendix A. Table 1 Panels A and B report summary statistics for the main variables used in the site-visits and forecast-errors analyses, respectively. The mean sample firm in the site-visits sample (Panel A) has book value assets of 18.8 billion RMB ($2.6 billion), a book leverage ratio of 43.1%, and an average ROA of 7.0%, which are comparable to those reported by Liu et al. (2017). In comparison, firms in the forecast-errors sample (Panel B) are larger (23.4 billion vs. 18.8 billion RMB). However, analyst coverage is comparable between the two samples (16.4 vs. 15.4).

Table 1 Panel C presents the summary statistics for the three variables we employ to capture the impact of analyst research on stock price efficiency: market reaction to analyst forecast revisions (*FRINFO*), market reaction to analyst recommendations, and market reaction to earnings announcements (*EAINFO*). We define a forecast revision to be any current-year forecast issued by an analyst team after the prior-year earnings announcement but before the current-year earnings announcement. We exclude revisions issued on the same day when earnings are announced, because attributing market reaction to analyst revisions rather than earnings announcements is difficult. We collect all annual earnings-announcement dates from CSMAR. Market reaction to analyst -forecast revisions, *FRINFO*, is measured as the three-day (-1,+1) size-adjusted cumulative abnormal return surrounding analyst-forecast revisions (Frankel et al. 2006), if the earnings-forecast revision is greater than or equal to 0. If the earnings-forecast revision is negative, we multiply it by –1. *FRINFO* has a skewed distribution with a mean of 0.916% and a median of 0.444%.

We obtain analyst stock recommendations from CSMAR and validated the quality of these data detailed in Internet Appendix IA.2. We measure the content of the analyst recommendation information through its price impact, as follows. First, we calculate the buy-and-hold return over the trading period (-1, +1), subtracting the corresponding buy-and-hold value-weighted market return for the same period, where 0 represents the recommendation issuance date. Second, following Jia et al. (2017), we classify recommendations into three categories: revisions, initiations, and reiterations. That is, if a revision is made within one year, either upgrade or downgrade, it is considered a revision; if no recommendation is issued within one year preceding the current recommendation, it is deemed an initiation; if an analyst reiterates her recommendation within one year, the current recommendation is considered a reiteration. In China, the issuance of reiterations is associated with a significant market reaction and thus contains information.[9] We therefore include reiterations of recommendations in our analysis. Third, for recommendation revisions, we "invest" $1 in the firm stock with an upgrade and short $1 with a downgrade. For recommendation initiations or reiterations, we invest $1 for "strong buy" or "buy," 0$ for "neutral," and short one dollar for "strong sell" or "sell." Last, we multiply the buy-and-hold return with the investment amount to obtain the market-reaction measure. We exclude observations with stock recommendations issued on the same day as the earnings announcement. The market reaction to stock recommendations also has a skewed distribution with a mean of 1.212 % and median of 0.407%. Thus, over our sample period, an average analyst recommendation appears to have information value.

We measure the market reaction to earning announcements as absolute three-day size-adjusted cumulative abnormal returns around the annual earnings announcement (*EAINFO*), following Merkley

[9] Reiterating recommendations represent over 60% of both our sample and the CSMAR recommendation universe. In Internet Appendix IA.2 Panel B, we show the three-day market reaction to different types of recommendations, consisting of reiterations within one year, those longer than one year, initiations without any prior recommendation, and upgrades and downgrades within one year and longer than one year. Indeed, the evidence suggests reiterations within one year contain value-relevant information—the market reaction is positive and statistically significant.

et al. (2017) and Grennan and Michaely (2020). The mean and median market reaction to the annual earnings announcement (*EAINFO*) are 3.805% and 2.752%, respectively, with a standard deviation of 4.500%.

3. Empirical Results

3.1 Cost of information and information production

Because HSR introductions are staggered over time, we are able to use a difference-in-differences approach to test the effect of information costs on information collection, by comparing the number of analyst visits before and after HSR introduction between the treated and control sample, as follows:

$$y_{a,i,t} = \beta_0 + \beta_1 HSR_{a,i,t} + Controls_t + \alpha_{a,i} + \alpha_{city(i)} \times \alpha_t + \alpha_{city(a)} \times \alpha_t + \varepsilon_{a,i,t}, \quad (1)$$

where a indexes brokerages, i indexes firms, t indexes years, *city(i)* and *city(a)* index the city where firm *i* and brokerage *a* are located, respectively, and $y_{a,i,t}$ is the number of analyst site visits (in logs) in year t. Equation (1) is also employed to estimate the HSR effect on other outcome variables comprising forecast errors, and market reaction to forecast revisions and recommendations.[10] As defined in the previous section, $HSR_{a,i,t}$ takes the value of 1 for year t if at least one HSR connects firm *i*'s city and broker *a*'s city at the beginning of t. We lag HSR introduction by one year to allow its effect to shore up. Controls include firm size, leverage, ROA, book-to-market, return volatility, the number of analysts following, and an analyst's firm-specific experience. All variables are defined in Appendix A. $\alpha_{a,i}$ are firm-broker-pair fixed effects. By including this set of fixed effects, we essentially hold a firm-broker pair constant and explore the change in site visits by the same broker for the same firm, in response to the information shock—HSR introduction. We include city-by-year fixed effects ($\alpha_{city} \times \alpha_t$) with respect to firm i's city and broker a's city. These two sets of fixed effects are meant to further control for shocks to firm i's city and broker a's city that might lead to the introduction of HSR and simultaneously affect analyst site visits. ε is the error term. To minimize the effect of outliers, we

[10] As shown in Table 1, the distribution of all the outcome variables are positively skewed. In these cases, models with log-transformed variables are more likely to satisfy the Classic Linear Model assumptions than models using the original variable (Wooldridge 2002).We therefore log-transform all these variables in the empirical tests.

winsorize all continuous variables at the 1st and 99th percentiles. The main coefficient of interest, β_1, captures the change in broker site visits in response to HSR introduction between the broker's and firm's city. If HSR introduction reduces information acquisition costs, which in turn affects the amount of information collected, we expect β_1 to be positive. To allow for serial dependence of the error terms, we cluster standard errors at the firm level.

Table 2 column (1) presents results based on equation (1), where the number of site visits (in logs) serves as the dependent variable. To save space, we omit reporting the coefficient estimate on control variables, and the full set of results are reported in Appendix B column (1). The coefficient on HSR introduction (*HSR)* is positive and statistically significant, implying analysts travel to visit firms more often after HSR introduction. The coefficient, 0.049, indicates treated analyst teams, relative to controls, increase their number of site visits by 4.9% post HSR introduction. In addition, Appendix B column (1) shows that analysts tend to visit a firm less when they have more firm-specific experience and thus more knowledge of the firm.

We also examine whether firm- and broker-specific shocks explain our findings. For example, if firms adopt a new policy of attracting institutional investor ownership, such a policy might increase the demand for analysts' research, and thus brokerages might lobby for an HSR connection with the firm. In this case, the firm's new policy represents a correlated omitted variable, leading to both HSR introduction and the increase in analysts' information production. To address this issue, we control for broker and firm time-varying fixed effects. As a result, firm-level characteristics and city-by-year fixed effects are subsumed in estimation. Results presented in column (2) show the coefficient on *HSR* continues to be positive and significant. Thus, the results of site-visits analysis are robust to controlling for analyst and firm time-varying effects. Importantly, the coefficient magnitude weakens, compared to that in column (1), underscoring the importance of explicitly considering time-varying shocks to brokers and firms, which prior studies cannot do (Giroud 2013; Bernstein et al. 2016).

We further stress-test our findings by examining how the impact of HSR on information collection varies with the difficulty of an analyst's visit to the firm. If indeed the HSR effect is due to reductions in information costs, we would expect the results to be more pronounced among analysts with greater difficulty in visiting the firm prior to HSR introduction. We categorize a firm as difficult to visit if the distance between analyst's location and the firm is greater than 800 km (500 miles) and there is no direct airline route between the two cities.[11] Accordingly, we code a dummy variable, *Difficult to visit*, as taking a value of 1 if the firm is categorized as difficult to visit in 2005 (predating the first HSR line). Roughly 47% of observations are classified as difficult for analysts to visit for both samples (Table 1 Panel A row 10 and Panel B row 11). We interact *Difficult to visit* with *HSR* in equation (1), and report the results in column (3). The coefficient on the interaction term, *HSR* Difficult to visit*, is positive and statistically significant, indicating the HSR effect is more pronounced for firms that are difficult for analysts to visit. From an economic perspective, the HSR effect on site visits is close to twice as large for firms that are difficult for analysts to visit as for other firms.

As a placebo test, we use firms located in Beijing, Shanghai, Guangzhou, and Shenzhen, together with analysts who are located in the same city as the covered firm, both of which we exclude from our main sample. Our hypothesis suggests HSR introduction should not affect analysts' visits to firms in this sample, because ease of travel has been minimally affected. Based on the results reported in Internet Appendix IA.3 column (1), we find an insignificant coefficient for *HSR* (-0.025), consistent with our argument that reductions in information-acquisition costs are responsible for the HSR effect on analysts' site visits.

Finally, we examine whether lobbying efforts (that might be correlated with pair-cities growth for example), can explain our findings. Suppose, for example, the economic link between Wuhan (firm-city) and Guangzhou (broker-city) strengthens, which induces the pair-city governments to lobby for introduction of a 1,000 km HSR connecting the two cities (and as a by-product, the HSR also connects other cities along the line such as Changsha and Yingde). At the same time, investors' information demand for Wuhan firms grows, resulting in higher analyst information production. Thus, finding a

[11] Research demonstrates that HSR has a greater impact on travel for distances above 500 km (Wan et al. 2016), or 800 km (Chen 2017). Using 500 km, 600 km, or 700 km as the cutoff does not change any of our results.

positive relation between HSR introduction and analyst information production could in part be an outcome of omitted shocks to a pair-city (of Wuhan and Guangzhou in the example above). We evaluate this possibility using two distinct experiments. First, we exclude all firm-broker-pair cities located on the two ends of an HSR (5.3% sample observations), because any pair-city government is unlikely to be able to successfully lobby for the introduction of a new HSR with the two ends elsewhere, that is, an HSR line where the firm-broker-pair cities are not at either end of the line. Our results reported in column (1) of Internet Appendix IA. 4 Panel A continue to hold: the coefficient on *HSR* is 0.045 and statistically significant. The second experiment is predicated on the notion that larger firms have greater incentives and more resources to lobby successfully (e.g., Zingales 2017) for HSR introductions. Thus, to further address the possibility of endogenous HSR introductions due to firm lobbying effort, we exclude the top 5% of firms based on total assets. Our results tabulated in column (1) of Internet Appendix IA.4 Panel B are robust to this reduced sample (the coefficient on *HSR* is 0.051 and statistically significant).

Overall, our findings in Table 2 establish that with easier access to portfolio firms, analysts visit more frequently, consistent with the notion that a lower cost of gathering information leads to greater information collection. The inclusion of broker and firm time-varying fixed effects and the consideration of HSRs that do not involve firm-broker-pair cities on the two ends, as well as exclusion of large firms, suggest our results are unlikely due to endogenous HSR connections between firm-broker-pair cities. If greater collection improves information precision, this effect should be reflected in the quality of analysts' output. Our next set of analyses tests this prediction by examining the impact of HSR introduction on the quality of the output.

3.2 Cost of information and the quality of information production

Perhaps analysts' most important output is their earnings forecasts. We use the accuracy of their forecasts as an indicator for the quality of information they produce. To assess the effect of lower information costs (HSR introduction) on forecast quality, we again use the difference-in-differences approach (equation (1)), with earnings-forecast errors as the dependent variable. Because earnings forecasts are issued throughout the year, we also use one additional control variable, the natural log of forecast duration, to account for the positive relation between forecast duration and forecast errors documented in prior studies (e.g., Jackson 2005). The unit of analysis is at the individual earnings-forecast level.[12]

The first column of Table 3 shows HSR introduction significantly improves forecast quality, as indicated by the negative and statistically significant coefficient on *HSR*. This coefficient, -0.021, implies earnings-forecast errors decline by 2.1% after the treated analysts are connected to portfolio firms by HSR introduction. To assess the economic significance of the HSR effect, we benchmark it against stock return volatility, since this variable is known to affect analysts' forecast error (Joos et al. 2016). As shown in Appendix B column (2), the coefficient on stock return volatility is positive and significant. A one standard deviation increase in stock return volatility is associated with a 3.4% (=0.056*61.6%) increase in forecast error. Benchmarked against the impact of volatility, the 2.1% reduction in forecast error associated with HSR introduction is economically significant and within a reasonable range. The coefficient on other control variables reported in Appendix B column (2) is largely of the expected sign. For example, firms with higher ROA and covered by more analysts enjoy more accurate forecasts. When we control for broker and firm time-varying fixed effects, column (2) shows the coefficient on *HSR* continues to be negative and its magnitude in absolute value weakens compared to column (1), similar to the findings in the site visits analysis reported in Table 2.

We further stress-test our findings by examining whether the impact of HSR on information quality is stronger when analysts have difficulty visiting the firm. Finding supportive evidence would suggest the HSR effect on information quality likely operates through its impact on analyst information collection during site visits. Column (3) shows a negative (-0.031) and significant coefficient on *HSR* Difficult to visit*. This finding implies that, post HSR introduction, analyst forecast errors for difficult-to-visit firms decline four times as much as for other firms, captured by the coefficient on *HSR*, -0.009.

[12] Using one-year-ahead forecasts (which are less common than current-year forecasts) produces qualitatively similar results as our main findings. Results are reported in Internet Appendix IA.5.

Moreover, using firms located in Beijing, Shanghai, Guangzhou, and Shenzhen, along with analysts who are located in the same city as the firm, both of which we exclude from our main sample, we again find an insignificant coefficient for *HSR* (-0.006), reported in Internet Appendix IA.3 column (2). Similar to the site-visits analysis, we address endogenous HSR introduction due to pair-city government lobbying effort. As before, our results (reported in column (2) of Internet Appendix IA.4, Panel A, and column (2) of Panel B) suggest that lobbying efforts do not drive our results.

To verify that our findings of post-HSR differences between treated and nontreated analysts do not reflect a trend of persistent and progressively widening differences that started before HSR, we test the parallel time-trend assumption. To test for a possible pre-existing differential trend in site visits or forecast errors, we estimate a modified version of equation (1), where we allow the effect of HSR to vary by year in a five-year window surrounding HSR introduction. The yearly estimates of *HSR* are reported in Figure 2 and Appendix C, suggesting the parallel-trend assumption holds. We do not observe any difference in analyst site visits or forecast errors between treated and nontreated firm-brokerage pairs in the year before (t-1) and the year of HSR introduction (t). The difference shores up in the year after the treatment (t+1). We also find a significant coefficient on HSR^{2+} which captures the HSR effect two years after introduction in column (1), implying that HSR introduction has a long-run effect on site visits. In column (2), the coefficient on HSR^{2+} is -0.021. Though it is statistically insignificant at conventional level, the magnitude is comparable to that for HSR^{+1} (-0.029), suggesting that HSR introduction also produces a long-run impact on information quality.

Thus far, our evidence shows the reduction in information costs has a pronounced effect on both the amount of information collected and its quality: easier and faster travel between analysts and portfolio firms results in more frequent visits and higher information quality. Our interpretation is further supported by the finding that the effect of information accessibility on both information collection (site visits) and the quality of information is stronger when HSR introduction is expected to have a larger impact, that is, when travel time is affected the most.

These findings raise two important questions. First, do more frequent visits and improved quality of information production by analysts lead to improved information efficiency? Specifically, what is the impact on price formation? Addressing this question has direct implications for the impact of reduced information frictions on price efficiency, and for firms' cost of capital. Second, what is the nature of the information that analysts gather while visiting firms? Both issues are also relevant to the current debate concerning the role of analysts in financial markets—whether they produce information of value and whether they contribute to better functioning markets (e.g., Hong et al. 2010; Merkley et al. 2017). These issues are particularly relevant at the time when big data are increasingly available and declining in cost. If big data improve information quality about firm fundamentals (Farboodi and Veldkamp 2019) and the information that analysts collect during site visits is largely objective and quantitative, analysts might replace site visits with data or investors might replace analysts with big-data vendors (Grennan and Michaely 2020; Mihet and Philippon 2019). We shed light on these issues in the next two sections.

4. Information production and price formation

One of the unique features of our study is that we can observe the impact of lower costs of information acquisition and increased quality of information production on price formation. Specifically, we expect information to be impounded in prices faster. This implication is important. Although this expectation is common to any study investigating improvements in information environment, testing it empirically is not always feasible, often simply because a public equity market does not exist (e.g., Giroud 2013; Bernstein et al. 2016). Empirically, we expect a more pronounced stock-market response to information-release events by analysts (forecast revisions and stock recommendations). And, as more information produced and distributed by analysts is impounded in prices, we expect the reaction to firms' earnings announcements to contain less new information, implying a lower price reaction. Prior research makes similar arguments, saying, for example, that market reaction to quarterly earnings announcements is stronger when fewer analysts follow an industry (Merkley et al. 2017).

We re-estimate equation (1), with stock-market reaction to analyst forecast revisions, *FRINFO*, as the dependent variable (in logs). Because HSR introduction improves analyst forecast precision, we expect investors' reaction to forecast revisions to be stronger post HSR. Table 4 column (1) shows a positive and statistically significant coefficient on *HSR* after controlling for variables associated with the information content of analyst forecast revisions (e.g., Frankel et al. 2006). From an economic perspective, HSR introduction increases market reaction to analyst forecast revisions by 1.7%. Thus, investors seem to recognize the better quality of information that analysts provide, as HSR introduction leads to more frequent site visits. The more pronounced market reaction after the exogenous shock to information costs also suggests more information is incorporated faster into prices, thus improving price discovery.

Next, we estimate equation (1), with market reaction to analyst stock recommendations as the dependent variable (in logs). Consistent with our expectation that HSR introduction improves price efficiency, the coefficient on *HSR*, shown in Table 4 column (2), is positive and significant at the 1% level. Economically, HSR introduction increases market reaction to analysts' recommendations by 1.9%. Taken together, our evidence suggests lower information costs result in prices impounding information faster, likely contributing to greater price efficiency. The implication of better information is not only for the more pronounced market reaction to the information that analysts provide, but also for the lower market reaction to earnings announcements.

To examine how exogenous shocks to information-acquisition costs (HSR introduction) affect market reaction to earnings announcements, we need to develop a firm-level measure of the intensity of HSR connections easing analyst commuting to firms. We construct this measure in two ways: one is based on the number of analysts who are HSR connected to the portfolio firm (*PHSR1)* and the other is based on their research output, the number of research reports signed by HSR-connected analysts (*PHSR2)*. For each firm-year in the analyst-forecasts sample, we calculate *PHSR1* as the proportion of analysts who are connected to firm i by HSR at the beginning of year t. For example, if in 2011, a firm is covered by eight analysts and two of them are directly connected to the firm by HSR at the beginning of 2011, this ratio is 0.25 for that year. Similarly, *PHSR2* is calculated as the proportion of research reports signed by HSR-connected analysts, relative to the total number of research reports issued by all analysts covering firm i in t. Table 1 Panel C shows the mean of *PHSR1* and *PHSR2* is 0.191 and 0.192, respectively. We estimate the following model:

$$EAINFO_{i,t} = \beta_0 + \beta_1 PHSR1_{i,t}(PHSR2_{i,t}) + Controls_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}, (2)$$

where *i* and *t* index for firm i and year t, respectively. The dependent variable is absolute three-day cumulative size-adjusted abnormal returns to the annual earnings announcement (in logs). Our variable of interest is *PHSR1* and *PHSR2*, respectively. If HSR introduction allows more information to be impounded in the stock price faster, we expect a negative coefficient on this variable. To be parsimonious, we include a similar set of control variables as equation (1), (excluding analyst-firm experience, which is measured at the firm-analyst level and the unit of analysis here is at the firm level). The detailed definition of control variables is in Appendix A. α *and* δ represent firm and year fixed effects. Note the inclusion of firm fixed effects allows us to interpret the coefficient on our variable of interest as the effect of a within-firm change in the intensity of analyst HSR connections on the information content of earnings announcements.

Table 5 reports the results of this analysis. In column (1), the coefficient on *PHSR1* (the proportion of analysts who are connected to the firm by an HSR at the beginning of t) is negative and significant, which is consistent with our expectation that HSR connections reduce market reaction to earnings announcement as they allow more information to be gathered and distributed by connected analysts. A one-standard-deviation increase in *PHSR1* (0.343, Table 1 Panel C) is associated with a 1.3% (0.343*0.039*100%) reduction in market reaction to the earnings announcement (*EAINFO*). We find similar results in column (2), when *PHSR2* (the proportion of research reports signed by HSR-connected analysts relative to the total number of research reports issued by all analysts covering firm i in t) serves as the measure of the intensity of analyst HSR connections.

To summarize, prices become more efficient as analyst access to information becomes easier, with the introduction of HSR: The market response to information releases by analysts, be it earnings revisions or recommendations, is more pronounced. At the same time, the response to firms' earnings announcements is less distinct, suggesting analysts' easier access to information increases the speed at which information in released and the speed at which this information is incorporated into prices.

Our difference-in-differences estimation (in reduced form) conducted previously provides unbiased estimate of the impact of information acquisition costs on output quality. A related, yet unanswered, question is the direct impact of site visits on output quality. Addressing this issue allows us to evaluate the importance of information collection through site visits on informational efficiency. Because site visit is a choice variable, and thus endogenous, we instrument site visits with HSR introduction. In the first stage, we regress site visits on HSR introduction using the specification of equation (1). In the second stage, we regress (i) forecast errors, (ii) market reaction to analyst forecast revisions, and (iii) market reaction to recommendations on the predicted value of site visits obtained from the first stage along with other variables in equation (2). As a result of the two-stage joint estimation that requires site visits data available, the sample size is reduced significantly.[13]

Table 6 presents the results of 2SLS estimation. Though the sample size varies across the three tests, the coefficient estimate for *HSR* in the first stage is similar ranging between 0.055 and 0.066. Thus for the sake of brevity we only report the first stage results for which analyst forecast error serves as the dependent variable in the second-stage. As shown in column (1), the coefficient on *HSR* is positive and significant, consistent with that reported in Table 2, suggesting that HSR introduction increases analyst visits to their portfolio firms.

Next, we evaluate the results from the second stage regressions. Column (2) shows that the coefficient on *Predicted # of Analyst Site Visits)* is -0.359, implying a 1% increase in site visits results in a 0.36% reduction in forecast error. The results suggest that corporate site visits improve analyst information quality significantly. Moving to columns (3) and (4), we find the coefficient on *Predicted # of Analyst Site Visits)* is 0.456 and 0.409, respectively, suggesting that site visits advance price discovery significantly: a 1% increase in site visits is associated with a 0.46% increase in market

[13] Notably there are two key differences between the 2SLS and the reduced form estimation. First, the sample size is significantly smaller in 2SLS because we constrain the sample to firms traded at SZSE only, since site visits data are not available for firms traded at SHSE. Second, to enable the second stage estimation, the unit of observation in the first stage regression is at the broker-firm-forecast or broker-firm recommendation level depending on the outcome variable, which differs from the unit of observation at the broker-firm-year level used in the reduced form estimation (Table 2). As such the coefficient estimate on *HSR* in the first stage is slightly different from that reported in Table 2.

reaction to forecast revisions and 0.41% increase in analyst recommendation profitability. In summary, using 2SLS estimation, we are able to assess the impact of analyst site visits on the quality of information they collect and price efficiency. These results complement the reduced form estimation, taken together shedding light on the relation between acquisition costs, information collection, and informational efficiency.

Finally, the more efficient prices along with more accurate forecasts associated with the lower information costs resulting from the HSR introduction, raises two interesting issues. The first concerns brokers' and analysts' incentives to collect information. Does better information and more accurate forecasts affect brokers' revenue and analysts' compensation and career path? The second, related issue, is whether the significant increase in forecast accuracy in the post HSR era suggests that analysts should have also collected more information before the HSR introduction. Ideally, we would have liked to obtain direct information about brokers' revenue from analysts' activities, and analysts' compensation. But both are not observable. Neither is the change in these variables as a function of the quality of analyst forecasts and recommendations. Therefore we estimate the benefits to brokers and analysts from the additional information gathered using data on abnormal trading volume associated with analysts' forecasts, together with some necessary assumptions about the structure of brokers' revenue and analysts' compensation. As detailed in Appendix D, results suggest that the more accurate information produced by analysts in the post HSR period is associated with an increase of analysts' compensation of about 3% along with better career prospects. With some additional assumptions, back of the envelope calculations do not indicate suboptimal information collection (from the perspective of both brokers and analysts) prior to HSR introduction.

5. Information collection: Soft versus hard information

Although we find evidence that HSR introduction eases analyst access to information, understanding the type of information that analysts collect during their visits is important. Specifically, we are interested in learning whether site visits assist analysts in gathering information that can be classified as hard or soft. Hard information is deemed to be easily transmitted and processed and can mainly be reduced to numbers, whereas soft information is subjective, and its collection is contextual and difficult to be separated from its use and thus difficult to be summarized in a number (Liberti and Petersen 2019).[14] Understanding the type of information that analysts collect during site visits is not only relevant to analyst production per se, but also has implications for how financial technology might replace human input (Grennan and Michaely 2020). If a significant portion of the information that analysts collect during site visits is soft, intangible information, technology might not be able to easily replicate this type of information collection and processing. We conduct our investigation using two distinct and independent approaches. First, we empirically examine whether the HSR effect on analyst site visits and forecast errors varies with the importance of soft information in forecasting future performance. Second, we conduct a large-scale survey of financial analysts, asking them about the type of information they collect during site visits.

5.1 Soft versus hard information

The first step is to construct a measure of the importance of soft information that is independent of our exogenous shock to information production. Research argues and shows collecting soft information at a distance is costly (e.g., Petersen and Rajan 2002; Liberti and Petersen 2019) and that forecasts issued by local analysts are more accurate than forecasts issued by nonlocals (Malloy 2005). Building on the insight that distance mainly affects the quality of soft information, we construct a soft-information measure that exploits within-firm differences in analyst forecast accuracy that is associated with geographical distance. In essence, because distance has significant effect on analysts' ability to read and analyze soft information, if the forecasts issued by nearby analysts are more accurate than the forecasts of more distant analysts, the soft information is likely to matter for that firm. We then categorize this firm as one for which soft information is important.

To calculate the measure, we first partition analysts following a firm into a distant and near group, based on the median distance between the analyst and firm. We then calculate the average forecast accuracy for each analyst group in the past three years and compare the forecast accuracy

[14] One can further argue that some inside information is also more likely to be transmitted through face-to-face interaction and hence can fit in the definition of soft information, despite Chinese regulations similar to Reg FD that prohibit it.

between the two groups. The three-year requirement is to smooth year-specific shocks (e.g., financial crisis of 2008) that might cause abnormal forecast errors. Finally, we code *Soft information* equal to 1 if earnings forecasts issued by brokers near the firm are more accurate than earnings forecasts issued by distant brokers, and 0 otherwise. We test the robustness of our classification of soft and hard information in Appendix E. To construct the measure, we require a firm be covered by at least two analysts located in different cities through year t-3 to t-1. The data requirement results in reduction of the sample size: the site-visits sample reduces to 28,309, from 33,200, and the forecasts sample reduces to 137,558, from 161,785.

As shown in Table 1 Panel A (row 11) and Panel B (row 12), roughly 52% of our sample firms are classified as firms where soft information is important (*soft information* =1). Untabulated results show, when *Soft Information* takes the value of 1, the mean forecast error is 1.054 for near analysts, much lower than the corresponding mean of distant analysts, 1.506, and the difference between the two is statistically significant. Interestingly, the opposite is observed when *Soft information* takes the value of 0: the means are 1.553 and 1.108, respectively, for near and distant analysts, and the difference between the two is also statistically significant. This observation is consistent with our argument that distance matters for the production of soft information.

We test whether the HSR effect varies with the importance of soft information by interacting *HSR* with *Soft Information* in equation (1). Because the soft-information measure does not vary within a firm, it is subsumed by the firm-broker-pair fixed effects. Table 7 columns (1), in which site visit serves as the dependent variable, shows the coefficient on the interaction term, *HSR*Soft information,* is positive and statistically significant, implying HSR introduction affects analyst information gathering for firms for which soft information matters. By contrast, *HSR* is not significantly associated with site visits, suggesting HSR introduction has little impact on analyst site visits when soft information is less relevant for forecasting performance. These results highlight that soft information plays an important role in analysts' decisions to make visits. If hard information was a key factor, we would have observed a significant coefficient on *HSR* for firms whose soft information is less important. Compared to firms for which soft information is less important, the HSR effect among their counterparts for which soft information is important almost doubles (0.069 (=0.036+0.033) vs. 0.036).

Likewise, we examine whether the HSR effect on forecast errors varies with the importance of soft information, and the results are presented in column (3). Consistent with the result in column (1), *HSR*Soft information* is negative and statistically significant, but the coefficient on *HSR* is insignificant. Moving from a firm whose soft information is less important to a firm whose soft information is important, the effect of HSR introduction on forecast errors goes from –0.005 to –0.030 (= –0.005 – 0.025), a fivefold increase in absolute magnitude. Importantly, our measure of soft information suggests not only the change in the ease of travel affects analysts' visits and the accuracy of their forecasts, but also the combination of the importance of the distance (the difference in forecast errors between distant and near analysts) and the distance itself.

Research suggests that firms with high stock return volatility tend to be more informationally opaque (i.e., lack of reliable hard information) and thus hard to value (e.g., Leary and Roberts 2010). Accordingly we use return volatility as another proxy for the importance of soft information. The results reported in columns 2 and 4 are largely consistent with the notion that HSR introduction has stronger impact when soft information plays a larger role. Specifically, the positive coefficient on the interaction of *HSR* with *High return volatility* in column (2) is suggestive of a greater impact of HSR introduction on site visits among firms for which soft information matters more; and at the same time we also find a stronger HSR effect on forecast errors among these firms, as evidenced by the negative and significant coefficient on this interaction term in column (4). Overall, our findings confirm that the additional information gathered by analysts after the introduction of HSR, likely soft in nature, has a higher marginal value for informationally opaque firms.

Our evidence highlights that the benefits of infrastructure that eases information access mainly accrue to firms whose soft information is more value relevant. To the extent that big-data analysis improves the quality of information that is objective and quantitative, financial analysts still have an advantage when they collect soft information. This result might suggest the direction in which the job of financial analysts evolves: It may focus more on the soft information, whereas big data will concentrate on quantitative information.

5.2 From the horse's mouth: Surveys of financial analysts

To provide further insights into what analysts do and what type of information they collect during site visits, we supplement our empirical results with survey evidence. As an important byproduct, the survey evidence also allows us to analyze, from a different perspective, the impact of information costs on agents' collection of information and output quality. Namely, it allows us to further stress-test our empirical findings of a causal effect of HSR introduction on analyst site visits and their information quality and further address endogeneity concerns.

5.2.1 Survey design

We developed a questionnaire based on prior research (Brav et al. 2005; Cheng et al. 2016; Bernstein et al. 2016). For all questions, participants were asked to state their degree of agreement with various statements concerning the effect of HSR, using a five-point scale (1 = *strongly disagree*, 5 = *strongly agree*). We then administered the questionnaire to eight analysts to verify the appropriateness of the terminology used and to ensure the instructions are unambiguous. The final versions of the questionnaire were formally distributed to 495 survey participants via a survey protocol (Wenjuanxing).[15]

We distributed the survey among two groups of analysts. The first group was composed of analysts from remote areas (defined as cities other than Beijing, Shanghai, Shenzhen, and Guangzhou) who had more difficulty visiting portfolio firms, and the second group included analysts from the four large metro areas who had less difficulty visiting, even before the HSR introduction. This separation enables us to compare and contrast the survey evidence between the two groups and further examine the effect of the change in degree of difficulty for analysts to visit a firm on information collection and production. In total, 202 "remote" analysts and 293 mega-city analysts received the survey. 98 remote

[15] Wenjuanxing is a widely used online survey platform in China that allows the distribution and collection of surveys through mobile devices and social platforms. Responses are automatically collected upon respondents' submissions. The online system can prevent invalid responses and improve efficiency.

and 236 mega-city analysts responded, respectively. The response rate (especially from the Beijing analysts) is high compared to prior studies (e.g., Brav et al. 2005), because we use the alumni of the University of International Business and Economics (UIBE), with which one of the coauthors is affiliated.[16]

The geographical distribution of the survey respondents is presented in Table 8 Panel A. We compare analysts and their portfolio firm characteristics between the remote (first row) and the mega-city groups (second row) in Table 8 Panel B. Roughly 7% of remote analysts cover more than eight firms, compared to 47% of the mega-city analysts. Nearly 80% of the remote group and 72% mega-city respondents are connected to at least one portfolio firm by HSR. Therefore, the two samples contain enough HSR-connected analysts, which allows us to analyze the HSR effect. 56% of the remote and 51% of the mega-city respondents, respectively, report that they pay one to three visits to each portfolio firm annually. For two reasons, both numbers are higher than the corresponding number for the sample used in the site visit analysis. First, the scale of the survey is too coarse to gauge analyst information-collection activities, compared with the empirical data (i.e., the survey gives no choice for site-visit frequency to lie between 0 and 1). Second, the empirical sample excludes but the survey sample includes local analysts. As is evident in the survey results, both remote and mega-city analysts visit local firms more frequently than nonlocal firms. Thus, survey evidence indicates analysts' site-visit decisions seem to be associated with geographical proximity.

5*.2.2 Survey results*

The responses to our survey questions are reported in Panels C and D of Table 8, based on remote and mega-city respondents, respectively. The full distribution of responses for the two groups is also depicted in Figure 3 Panels A and B. Our discussion focuses on Table 8 Panel C, the results for remote analysts, because they are presumably more responsive to HSR introduction. We begin with the first question of whether direct HSR connection results in analysts conducting more site visits in person. Ninety-six percent of respondents said yes, with a mean response of 4.59 out of five (Table 8, Panel C,

[16] Survey methodology is not without potential shortcomings. Two common concerns are social desirability bias and description variance bias. In online Appendix IA.6, we discuss them in details along with how we address the potential concerns.

Row 1), which is statistically significant from the neutral midpoint response of 3 at the 1% level.[17] We find a similar response (91.84%) when the same question is asked directly concerning analysts' own behavior: "The introduction of a direct HSR route will increase *your* frequency of visit" (Table 8, Panel C, Row 6). When we separate the responses to the positively worded survey (original) from the negatively worded one (alternative), the agreement rate is 89% for the former (mean response = 4.58) and 2.33% for the latter (mean response = 1.40), as indicated in Internet Appendix IA.9 Panels A and B.[18] Thus, whether questions are asked directly or indirectly, worded positively or negatively, survey evidence is consistent: HSR introduction increases analysts' inclination to visit firms more often. When we further asked whether HSR introduction increases analysts' flexibility to visit firms, 96% indicated HSR introduction reduces analysts' information-acquisition costs, due to increased travel flexibility. In sum, survey evidence corroborates our empirical findings that easier travel to firms, due to HSR introduction, allows analysts to visit more often.

93% of respondents believe (Row 2 in Table 8, Panel C) HSR introduction helps analysts obtain more firm-specific information. Although firm-specific information can be both hard and soft, our next set of questions centers on soft information. Roughly 87% respondents agreed a direct HSR connection results in analysts better understanding corporate strategies, operations, performance, key challenges/issues facing companies, corporate culture, and employee morale (Rows 3, 4, and 5 in Table 8, Panel C). Information of this sort likely involves personal assessment and depends on the context and thus can be considered soft. Overall, the survey evidence not only echoes our cross-sectional findings, but also yields additional insights into what type of information analysts collect during visits.

Finally, we asked whether HSR introduction helps them make more accurate earnings forecasts and better long-term growth forecasts. 82% of respondents told us HSR introduction helps them make more accurate earnings forecasts (Row 10 in Table 8, Panel C), and 78% said HSR introduction helps them forecast a company's long-term growth (Row 11 in Table 8, Panel C). Only 8% and 16%

[17] The mean responses to all survey questions are statistically significantly different from the neutral mid-point response of 3. For brevity, we omit discussing it for other survey questions.
[18] The responses to all our original survey questions are consistent with those to the alternative. To save space, we omit the discussion for all other questions.

respondents (Figure 3) thought HSR introduction does not help with forecasting earnings and long-term growth, respectively. Simple correlations indicate respondents who believe HSR introduction helps them increase site visits also tend to agree it will help them collect more soft information (correlation coefficient = 0.874, p-value <0.001), and make better forecasts of earnings (correlation coefficient = 0.654, p-value <0.001) and of long-term growth (correlation coefficient = 0.649, p-value <0.001). This result suggests analysts' increased site visits, due to HSR introduction, likely allow them to acquire more soft information, contributing to improved forecast accuracy. The evidence corroborates our empirical findings.

Moving to the mega-city respondents, we expect the HSR impact to be weaker, due to their easy access even without HSR. As shown in Panels D of Table 8, the agreement rate to all questions is indeed lower for mega-city respondents. For example, 87% mega-city respondents (Row 1 of Table 8 Panel D) versus 96% remote analysts (Row 1 of Table 8 Panel C) agree an HSR connection results in analysts conducting more site visits, and the difference between the two is statistically significant.[19] The difference in the responses between remote and mega-city groups thus corroborates our empirical findings that the HSR effect is more pronounced among firms that are difficult for their analysts to visit. The difference also suggests respondents do not merely respond to questionnaires mechanically. If they did, we would have observed similar responses between the two groups. Furthermore, our survey findings do not simply reflect respondents' personal experience with HSR, because we find no significant difference in responses between respondents with an HSR connection to their portfolio firms and those without one.

Though based on a completely different approach, our survey provides evidence entirely consistent with empirical findings that HSR introduction increases the frequency of analysts' visits to their portfolio companies and helps them make more accurate earnings forecasts. This consistency of evidence alleviates endogeneity concerns about the empirical analyses. Moreover, the survey reveals

[19] The difference in the agreement rate between remote and mega-city analysts for all questions is statistically significant.

analysts collect firm-specific information during site visits, particularly soft information that is hard to obtain otherwise.

6. Extensions: Extensive margin, External Validity, the Effect of Governance and Analysts Time Constraints

Our data enable us to explore several additional aspects of the impact of changes in information acquisition costs. First, an important question is whether the exogenous reduction in information costs also affects the extensive margin−the decision of whether a new firm will be covered. Second, having established the casual impact of reduction in information costs on analysts' information production, we examine the external validity of these results using the entire panel data. Third, differences in governance may affect the quality of information, especially soft information, firms provide to analysts and markets. We therefore examine how governance quality affects the change in information production (i.e. forecast error) associated with the HSR introduction.

6.1 The effect of HSR introduction on analysts' stock coverage decision

Empirical analyses thus far focus on the effect of changes in analysts' information environment on their forecast error, as well as the market reaction to forecast revisions and recommendations. By construction the tests require data on these variables also before the HSR, necessitating us to focus on intensive margins. With that said, if proximity facilitates information gathering, one would expect it matters for the decision to initiate coverage as well. Therefore, we conduct additional tests examining the effect of HSR introduction on the extensive margin of analyst information production – initiation of research coverage.

We start with all firms that are included in the CSMAR database and require firms be covered by at least one analyst in the prior year to ensure the existence of market demand for analyst research. For a broker-firm-year to be included in the coverage initiation sample we require it did not cover the firm (e.g., no earnings forecast) in the prior three years (t-3 up to t-1), and then either issues an earnings forecast or continues to issue no forecast in year t. Accordingly, $Initiation_{a,i,t}$ is set as a dummy variable, equal to 1 if broker *a* issues an earnings forecast for firm *i* in year *t*, and 0 otherwise. Using this sample, we test whether a broker's coverage decision in t is affected by the HSR connectivity between the broker-firm pair as of the beginning of t based on equation (3) below.

$$Initiation_{a,i,t} = \beta_0 + \beta_1 HSR_{a,i,t} + Controls_t + \alpha_{a,i} + \alpha_{city(i)} \times \alpha_t + \alpha_{(a)} \times \alpha_t + \varepsilon_{a,i,t}, \quad (3)$$

where *a* indexes brokerages, *i* indexes firms, *t* indexes years, *city(i)* index the city where firm *i* is located. $HSR_{a,i,t}$ takes the value of 1 for year t if at least one HSR connects firm *i*'s city and broker *a*'s city at the beginning of t. Controls include firm size, leverage, ROA, book-to-market, return volatility, the number of analysts following, and an analyst's firm-specific experience. All variables are defined in Appendix A. ε are the error terms. To allow for serial dependence of the error terms, we cluster standard errors at the firm level.

The results are reported in Table 9. Column (1) focuses on firms traded at both SHSE and SZSE. We find a positive and significant coefficient on *HSR*. HSR introduction increases the likelihood of a broker initiating research coverage for a connected firm by 0.6%. Compared to the sample mean *Initiation* of 4.2%, this 0.6% increase is economically significant. Similar evidence is observed in column (2) when we focus on firms traded at SZSE only. By reducing information acquisition costs, HSR introduction leads to a significant increase in analysts' initiation of research coverage for the stock that analysts are connected to. Combined with previous findings, our evidence suggests that reductions in acquisition costs lead to greater information production at both intensive margin and extensive margin.

6.2 Distance and analyst information production-External validity test

To evaluate the external validity of our results, we use the entire panel data to examine whether distance constraint to visit a firm is associated with analyst consensus forecast error and market reaction to firms' earnings announcements. We construct a firm-level proxy for distance constraint, *Average difficult to visit*, as the mean of *Difficult to visit* defined in section 3.1. Results reported in Appendix F show a positive and significant coefficient on *Average difficulty to visit* in both columns, consistent with the notion that distance has a negative impact on analyst information production and price efficiency. It is reassuring that the results of the panel data regression are consistent with our natural experiment based on HSR introduction, thereby enhancing the external validity of our study.

6.3 The effect of corporate governance and analyst time constraints

Our evidence indicates that information revealed through analysts' site visits leads to increased price efficiency. One can argue that managers may have the incentive to convey this soft information even before those site visits intensified following the introduction of HSR. To the extent that managers in weakly-governed firms might have fewer incentives to share information with analysts, for example, in order to protect their private benefits, the additional information gathered by analysts likely leads to a stronger HSR effect on their forecast errors among weakly-governed firms. We test this prediction using two measures of corporate governance. Predicated on the findings that government-owned enterprises are more prone to agency issues (e.g., La Porta, Silanes and Shleifer 2002), our first measure is whether a firm is a state-owned enterprise (*SOE*).We use board independence as our second measure of governance (e.g. Gompers, Ishii and Metrick 2003). As shown in Table 10, the coefficient on the two interaction terms, *HSR*SOE* and *HSR*Low board independence*, is both negative and significant, indicating a more pronounced HSR effect on analysts' forecast errors among weakly-governed firms. The evidence is consistent with the notion that poorly governed firms have fewer incentives to produce information, although doing so facilitates external monitoring.

The evidence above suggests the effect of HSR introduction intensifies when firms tend to provide less information. Our earlier evidence also indicates the HSR effect varies with analysts' geographical constraints (i.e., the difficulty to visit a portfolio firm). The next test analyzes how analysts' travel costs affect this effect. If travel time matters for analysts gathering information, we expect HSR introduction might have a bigger impact on analysts who are more time-constrained. We test this prediction using two proxies for time constraints–*High industry experience* (experienced analysts might cover a larger portfolio of firms, leaving them less time to travel to each individual firm, especially before the HSR introduction), and *Travel constraints* (the sum of the distances (i.e., the total distance) between each of the portfolio firms and the analyst's location). To conserve space, the results are reported in Appendix G. Consistent with our prediction, we find a positive coefficient for the interaction term between *HSR* and each of the two proxies of time constraints, although only the one with industry experience is statistically significant. The results further suggest that by lowering travel costs, HSR enables analysts to gather information more frequently.

7. Conclusion

This paper uses novel datasets of Chinse analysts' site visits, combined with a measure of the quality of their output, to empirically examine the effect of acquisition costs on information production, information quality, and the efficiency of outcomes. To further enhance our ability to draw causal inferences, we also use a large survey of 334 financial analysts. We first find HSR introduction leads to a significant increase in information production, measured by the frequency of analysts visiting portfolio firms. Survey evidence strongly supports this empirical finding: over 90% of respondents believe HSR introduction increases analysts' frequency of site visits. They also believe HSR introduction enhances analysts' flexibility of travelling to portfolio firms, further confirming HSR reduces information-acquisition costs, essentially leveling the playing field between distant and near analysts.

Second, we find evidence that reductions in information-acquisition costs improve the quality of analysts' outputs, measured by the accuracy of their earnings forecasts. Our survey results confirm this finding: roughly 82% analysts believe HSR introduction helps make better forecasts of earnings and long-term growth. Additionally, we find reductions in information-acquisitions costs also lead analysts to produce more information at the extensive margin─a higher likelihood of covering a new firm.

Third and importantly, we find analysts' information environment affects information efficiency. Market reaction to analyst forecast revisions and recommendations increases significantly, whereas reaction to firms' earnings announcements declines post HSR introduction. This finding suggests stock prices become more efficient as information promulgates in a timelier manner. This improved efficiency is important not only in its own right, but also because it offers evidence of the importance of sell-side analysts in price discovery: analysts bring real and tangible value to financial markets. Further, although outside the scope of our study, the findings likely have implications for analysts' role in monitoring (Chen et al. 2015). Because reductions in acquisitions costs lower information asymmetry between analysts and the portfolio firms, particularly for firms with weak corporate governance, we expect analysts' ability to monitor portfolio firms to improve, thereby resulting in more efficient investments and higher firm performance.

As well as being important in the context of analysts' behavior, these findings support one of the fundamental conjectures made by leading theories of asymmetric information: lower information costs lead to greater information production. Due to lack of data on agents' information production and the difficulty in measuring the quality of agents' information precisely, empirical studies that examine the link between information-acquisition costs and the efficiency of outcomes de facto assume lower information costs lead to greater information production (e.g., allocation of goods by traders in the product markets (Jensen 2007; Aker 2010), allocation of resources by investors in capital markets (e.g., Dávila and Parlatore 2018), managers' decisions about internal resource allocation (e.g., Duchin and Sosyura 2013; Giroud 2013), and allocation of resources when assets are not freely traded in the case of VCs (Bernstein et al. 2016), or regulatory oversight for commercial banks (Gopalan, Kalda, and Manela 2018)). To our knowledge, our study is the first to empirically test and validate the causal links between acquisition costs and information production, information quality, and the efficiency of outcomes (Grossman and Stiglitz 1980; Verrecchia 1982).

Finally, consistent with the notion that constraints on personal interactions matter more for soft information, we show the impact of HSR on information collection and accuracy is stronger among analysts for whom site visits are costlier without HSR and for firms for which soft information is more important. Survey evidence further corroborates these conclusions and highlights the critical role of soft-information collection in these site visits. Our findings thus suggest face-to-face interaction cannot be easily substituted by technology when soft information is important for the agent decision-making. Although crowd-sourced forecast aggregators provide incremental information, beyond traditional information providers (Da and Huang 2020), sell-side analysts' research shows significant substitution between investors' use of traditional information and financial technology (Grennan and Michaely 2020). Our findings underscore that sell-side analysts might be able to maintain their comparative advantage when human input is important, despite technological advances easing the collection and transmission of hard information (e.g., Mihet and Philippon 2019).

Policy makers at the macro level may also gain insights from our findings: when financial markets, like in China, are not fully developed, an investment in infrastructure (e.g., railways, highways, faster internet network, and faster and more efficient telecommunications), has a meaningful positive externality on informational market efficiency. While there is some macro-level evidence that construction of transportation infrastructure promotes economic growth (Duranton and Turner 2012;

Donaldson and Hornbeck 2016), the micro-level firm evidence has been limited, particularly in the context of financial markets. Thus, our paper fills the void by providing evidence that building infrastructure promotes informational efficiency in financial markets.

Reference

Agarwal, Sumit, Wenlan Qian, Amit Seru, and Jian Zhang, 2020, Disguised corruption: Evidence from consumer credit in China, *Journal of Financial Economics*, 137, 430−450.

Aker, Jenny C, 2010, Information from markets near and far: Mobile phones and agricultural markets in Niger, *American Economic Journal Applied Economics* 2, 46–59.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, 2016, The impact of venture capital monitoring, *Journal of Finance* 71, 1591–1622.

Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely, 2005, Payout policy in the 21st century, *Journal of Financial Economics* 77, 483−527.

Chae, Joon, 2005, Trading volume, information asymmetry, and timing information, *Journal of Finance* 60, 413−442.

Chen, Kailin, and Fuxin Jiang, 2017, The current issue of commission fee in Chinese securities industry, *Management*, 82−83.

Chen, Tao, Jarrad Harford, and Chen Lin, 2015. Do analysts matter for governance? Evidence from natural experiments, *Journal of Financial Economics* 115, 383−410.

Chen, Zhenhua, 2017. Impacts of high-speed rail on domestic air transportation in China. *Journal of Transport Geography* 62, 184−196.

Cheng, Qiang, Fei Du, Xin Wang, and Yutao Wang, 2016, Seeing is believing: Analysts' corporate site visits, *Review of Accounting Studies* 21, 1–42.

Da, Zhi, and Xing Huang, 2020, Harnessing the wisdom of crowds, Management Science 66 (5), 1847 –1867.

Dávila, Eduardo, Cecilia Parlatore, 2018, Identifying Price Informativeness, NBER Working Paper No. w25210.

Donaldson, Dave, and Richard Hornbeck, 2016, Railroads and American economic growth: A "market access" approach, *The Quarterly Journal of Economics* 131, 799−858.

Duchin, Ran, and Denis Sosyura, 2013, Divisional managers and internal capital markets. *Journal of Finance* 68, 387−429.

Duranton, Gilles, and Matthew A. Turner, 2012, Urban growth and transportation, *Review of Economic Studie*s 79, 1407−1440.

Farboodi, Maryam, and Laura Veldkamp, 2019 A Growth Model of the Data Economy, Working Paper, Columbia Business School, New York

Frankel, Richard, S. Kothari, and Joseph Weber, 2006, Determinants of informativeness of analyst research, *Journal of Accounting and Economics* 41, 29−54.

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

Gompers, Paul, Joy Ishii, and Andrew Metrick, 2003, Corporate governance and equity prices, *The quarterly journal of economics* 118, 107−156.

Gopalan, Yadav, Ankit Kalda, and Asaf Manela, 2018, Hub-and-spoke regulation and bank leverage, Working paper.

Grennan, Jillian, and Roni Michaely, 2020, Fintechs and the market for financial analysis, *Journal of Financial and Quantitative Analysis*, Forthcoming.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Hong, Harrison, and Jeffrey Kubik, 2003, Analyzing the analysts: Career concerns and biased earnings forecasts, *Journal of Finance* 58, 313–351.

Hong, Harrison, Jeffrey D. Kubik, and Amit Solomon, 2000, Security analysts' career concerns and herding of earnings forecasts, *The Rand Journal of Economics*, 121−144.

Hong, Dong, Harrison Hong, and Andrei Ungureanu, 2010, Word of mouth and gradual information diffusion in asset markets, Working paper.

Huang, Yi, Marco Pagano, and Ugo Panizza, 2020, Local Crowding-Out in China, *The Journal of Finance* 75, 2855−2898.

Jackson, Andrew R, 2005, Trade generation, reputation, and sell-side analysts, *Journal of Finance* 60, 673–717.

Jensen, Robert, 2007, The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector, *The Quarterly Journal of Economics* 122, 879–924.

Jensen, Michael C., and William H. Meckling, 1976, Theory of the firm: Managerial behavior, agency costs and ownership structure, *Journal of Financial Economics* 3, 305−360.

Jia, Chunxin, Yaping Wang, and Wei Xiong, 2017. Market segmentation and differential reactions of local and foreign investors to analyst recommendations, *Review of Financial Studies* 30, 2972–3008.

Joos, Peter, Joseph D. Piotroski, and Suraj Srinivasan, 2016. Can analysts assess fundamental risk and valuation uncertainty? An empirical analysis of scenario-based value estimates, *Journal of Financial Economics* 121, 645−663.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer, 2002, Government ownership of banks, *Journal of Finance* 57, 265-301

Leary, Mark T., and Michael R. Roberts, 2010. The pecking order, debt capacity, and information asymmetry, *Journal of Financial Economics* 95, 332−355.

Liberti, José María, and Mitchell A. Petersen, 2019, Information: Hard and soft, Review of *Corporate Finance Studies* 8, 1–41.

Lin, Yatang, 2017, Travel costs and urban specialization patterns: Evidence from China’s high speed railway system, *Journal of Urban Economics* 98, 8-123.

Liu, Laura Xiaolei, Haibing Shu, and KC John Wei, 2017, The impacts of political uncertainty on asset prices: Evidence from the Bo scandal in China. *Journal of Financial Economics* 125, 286−310.

Lo, Andrew W., and Jiang Wang, 2006, Trading volume: Implications of an intertemporal capital asset pricing model, *Journal of Finance* 61, 2805−2840.

Malloy, Christopher J., 2005, The geography of equity analysis, *Journal of Finance* 60, 719–755.

Merkley, Kenneth, Roni Michaely, Joseph Pacelli, 2017, Does the scope of the sell-side analyst industry matter? An examination of bias, accuracy, and information content of analyst reports, *Journal of Finance* 72, 1285−1334.

Michaely, Roni, and Wayne H. Shaw, 1994, The pricing of initial public offerings: Tests of adverse-selection and signaling theories, *Review of Financial Studies* 7, 279−319.

Mihet, Roxana, and Thomas Philippon, 2019, The Economics of Big Data and Artificial Intelligence, *Disruptive Innovation in Business and Finance in the Digital World (International Finance Review)* 20, 29−43.

Petersen, Mitchell A., and Raghuram G. Rajan, 2002, Does distance still matter? The information revolution in small business lending, *Journal of Finance* 57, 2533−2570.

Wan, Yulai, Hu-Koo Ha, Yuchiro Yoshida, and Anming Zhang, 2016. Airlines' reaction to high-speed rail entries: empirical study of the Northeast Asian market. *Transportation Research Part A: Policy and Practice* 94, 532–557.

Wooldridge, Jeffrey M, 2002, Econometric analysis of cross section and panel data, *MIT press, Cambridge, MA.*

Verrecchia, Robert E.,1982, Information Acquisition in a Noisy Rational Expectations Economy, *Econometrica* 50, 1415−1430.

Zingales, Luigi, 2017, Towards a Political Theory of the Firm, *Journal of Economic Perspectives* 31, 113−130.

Wu, 2014. Chinese securities companies. *John Wiley & Sons, Inc.*, Hoboken, New Jersey.

Table 1. Summary Statistics

This table provides summary statistics for variables used in site-visits analyses in Panel A, forecast-errors analyses in Panel B, and stock-price-efficiency analyses in Panel C. *HSR* is an indicator variable that takes the value of 1 for year t if at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. Other variable definitions are provided in Appendix A.

Panel A

Variable	# of Obs.	Mean	S.D.	Median
# of Analyst Site Visits	33200	0.590	0.964	0.000
HSR	33200	0.405	0.491	0.000
Total assets (Billion Yuan)	33200	18.814	36.038	6.606
Leverage	33200	0.431	0.188	0.423
ROA	33200	0.070	0.060	0.062
Book-to-market	33200	0.419	0.308	0.341
Return volatility	33200	0.119	0.066	0.109
# of analyst following	33200	15.397	8.516	15.000
Firm experience	33200	2.066	1.496	1.500
Difficult to visit	33200	0.472	0.499	0.000
Soft information	28309	0.520	0.500	1.000

Panel B

Variable	# of Obs.	Mean	S.D.	Median
Analyst Forecast Error (%)	161785	1.216	1.666	0.600
Coverage	656526	0.104	0.305	0.000
HSR	161785	0.390	0.488	0.000
Total assets (Billion Yuan)	161785	23.414	41.643	8.731
Leverage	161785	0.455	0.186	0.457
ROA	161785	0.074	0.058	0.065
Book-to-market	161785	0.434	0.322	0.348
Return volatility	161785	0.119	0.056	0.109
# of analyst following	161785	16.421	8.766	16.000
Firm experience	161785	2.147	1.547	1.500
Difficult to visit	161785	0.469	0.499	0.000
Soft information	137558	0.527	0.499	1.000

Panel C

Variable	# of Obs.	Mean	S.D.	Median
Market reaction to analyst forecast revisions (%) (FRINFO)	53201	0.916	5.301	0.444
Market reaction to stock recommendations (%)	52374	1.212	5.233	0.407
Market reaction to earnings announcements (%) (EAINFO)	21590	3.805	4.500	2.752
Proportion of HSR-connected analysts (PHSR1)	21590	0.191	0.343	0.000
Proportion of earnings forecasts issued by HSR-connected analysts (PHSR2)	21590	0.192	0.349	0.000

Table 2. HSR Introduction and Analyst Site Visits

This table examines the effect of HSR introduction between a firm-brokerage pair on analyst site visits. The unit of analysis is at the broker-firm pair by year level. Column (1) includes firm-broker-pair fixed effects and city-times-year fixed effects. Column (2) includes firm-broker-pair fixed effects, and firm-year and broker-year fixed effects. Column (3) explores cross-sectional variation based on column (1) model specification. *HSR* is an indicator variable that takes the value of 1 if at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. The dependent variable, *# of Analyst Site Visits*, is the natural logarithm of 1 plus the number of site visits conducted by a brokerage for a firm in year t. *Difficult to visit* is an indicator variable that equals 1 if the distance between the listed firm and the brokerage house is more than 800 km and if no direct flight exists between the firm and the analyst by 2005, and 0 otherwise. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	# of Analyst Site Visits		
	(1)	(2)	(3)
HSR	0.049***	0.028**	0.031*
	(0.014)	(0.014)	(0.016)
HSR*Difficult to visit			0.046*
			(0.023)
Control variables	Yes	Yes	Yes
Broker-firm pair FE	Yes	Yes	Yes
Firm city × Year FE	Yes	No	Yes
Firm × Year FE	No	Yes	No
Brokerage city × Year FE	Yes	No	Yes
Brokerage × Year FE	No	Yes	No
Observations	31,417	29,977	31,417
R-squared	0.551	0.673	0.551

Table 3. HSR Introduction and Analyst Forecast Errors

This table examines the effect of HSR introduction between a firm-brokerage pair on analyst forecast errors. The unit analysis is at the forecast level. Column (1) includes firm-broker-pair fixed effects and city-times-year fixed effects. Column (2) includes firm-broker-pair fixed effects, and firm-year and broker-year fixed effects. Column (3) explores cross-sectional variation based on column (1) model specification. *HSR* is an indicator variable that takes the value of 1 if there is at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. The dependent variable, *Analyst Forecast Error*, is the natural logarithm of 1 plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. *Difficult to visit* is an indicator variable that equals 1 if the distance between the listed firm and the brokerage house is more than 800 km and if no direct flight exists between the firm and the analyst by 2005, and 0 otherwise. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Analyst Forecast Error		
	(1)	(2)	(3)
HSR	-0.021**	-0.011*	-0.009
	(0.009)	(0.006)	(0.010)
HSR*Difficult to visit			-0.031**
			(0.014)
Control variables	Yes	Yes	Yes
Broker-firm pair FE	Yes	Yes	Yes
Firm city × Year FE	Yes	No	Yes
Firm × Year FE	No	Yes	No
Brokerage city × Year FE	Yes	No	Yes
Brokerage × Year FE	No	Yes	No
Observations	160,519	158,842	160,519
R-squared	0.612	0.774	0.612

Table 4. HSR Introduction and Market Reaction to Analyst Research Releases

This table examines the effect of HSR introduction on the market reaction to analyst research releases, consisting of forecast revisions and stock recommendations. *HSR* is an indicator variable that takes the value of 1 if at least one HSR connects the firm-broker pair at the beginning of *t*, and 0 otherwise. In column (1), the dependent variable, *FRINFO*, is market reaction to analyst forecast revisions. In column (2), the dependent variable is *Market Reaction to Stock Recommendations*. Both dependent variables are in log-transformed with a small constant added to the original variable before log-transformation. Control variables are defined in Appendix A. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	FRINFO	*Market Reaction to Stock Recommendations*
	(1)	(2)
HSR	0.017**	0.019***
	(0.008)	(0.007)
Total assets	0.008	0.007
	(0.007)	(0.006)
Leverage	-0.007	0.009
	(0.026)	(0.024)
ROA	0.165**	0.062
	(0.069)	(0.054)
Book-to-market	-0.045***	-0.075***
	(0.011)	(0.011)
# of analyst following	-0.010*	-0.008
	(0.005)	(0.005)
Firm experience	-0.000	-0.004
	(0.005)	(0.004)
Forecast duration	0.007*	
	(0.004)	
Broker-firm pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	47,783	46,948
R-squared	0.297	0.305

Table 5. HSR Introduction and Market Reaction to Firm Earnings Announcements

This table examines the relation between the intensity of analyst HSR connections and the information content of earnings announcements. The unit of analysis is at the firm-year level. Firm fixed effects are included in the estimation. The dependent variable, *EAINFO*, is the natural logarithm of one plus market reaction to earnings announcements, calculated as size-adjusted absolute cumulative abnormal returns for the three-day window (-1, +1) around the annual earnings announcement date and multiplied by 100. *PHSR1* is the proportion of analysts who are connected to the firm by an HSR at the beginning of t, relative to the total number of analysts covering the firm in year t. *PHSR2* is the proportion of research reports signed by HSR-connected analysts relative to the total number of research reports signed by all analysts covering the firm in t. An analyst is HSR connected if the analyst is connected by an HSR with the portfolio firm at the beginning of t. Control variables are defined in Appendix A. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	EAINFO	
	(1)	(2)
PHSR1	-0.039*	
	(0.021)	
PHSR2		-0.045**
		(0.021)
Total assets	-0.019*	-0.018*
	(0.010)	(0.010)
Leverage	-0.002***	-0.002***
	(0.000)	(0.000)
ROA	0.000***	0.000***
	(0.000)	(0.000)
Book-to-market	-0.053*	-0.053*
	(0.030)	(0.030)
# of analysts following	0.012	0.013
	(0.009)	(0.009)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Observations	21,533	21,533
R-squared	0.140	0.140

Table 6 The Elasticity of Outcomes to Site Visits- Two-stage Least Square Estimation

This table reports the results of two-stage least square estimation for the relation of analyst site visits with forecast error, market reaction to forecast revisions, and market reaction to analyst recommendations. The samples contain firms traded at SZSE only due to the availability of site visits data. The dependent variable in the first stage regression is the natural logarithm of one plus the number of analyst site visits in year t. Column (1) reports the first-stage results when analyst forecast error serves as the dependent variable in the second stage. Columns (2) – (4) report the results of the second-stage regression, where the dependent variable is *Analyst Forecast Error* , market reaction to analyst forecast revisions (*FRINFO)*, and market reaction to recommendations, respectively, with all in log transformation. The explanatory variable is the *Predicted # of Analyst Site Visits* obtained from the first-stage. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	First-stage		Second-stage	
	# of Analyst Site Visits	*Analyst Forecast Error)*	*FRINFO)*	*Market Reaction to Stock Recommendations*
	(1)	(2)	(3)	(4)
HSR	0.066***			
	(0.018)			
Predicted *# of* Analyst Site Visits		-0.359**	0.456*	0.409**
		(0.174)	(0.046)	(0.040)
Control variables	Yes	Yes	Yes	Yes
Broker-firm pair FE	Yes	Yes	Yes	Yes
Firm city × Year FE	Yes	Yes	Yes	Yes
Brokerage city × Year FE	Yes	Yes	Yes	Yes
Observations	79,366	79,366	23,356	22,931
R-squared	0.627	0.718	0.304	0.317

Table 7. The Effect of HSR Introduction and the Importance of Soft Information

This table examines the heterogeneous effect of HSR introduction on analyst site visits and forecast errors based on the importance of a firm's soft information for performance valuation. *HSR* is an indicator variable that takes the value of o1if at least one HSR connects the firm-broker pair at the beginning of *t*, and 0 otherwise. The importance of soft information has two proxies. The first is *Soft information*, an indicator variable taking the value of 1 if the average EPS forecast error for the near analyst group is lower than that for the distant analyst group for a given firm over year t-3 to year t-1. The second proxy is *High return volatility*, an indicator variable taking the value of 1 if the firm's stock return volatility is above the sample median in the year immediately before HSR introduction. In columns (1) and (3), the dependent variable, *# of Analyst Site Visits*, is the natural logarithm of 1 plus the number of site visits conducted by a brokerage for a firm in year t. In columns (2) and (4), the dependent variable, *Analyst Forecast Error*, is the natural logarithm of 1 plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. Standard errors clustered at the firm level are reported in parentheses. The superscripts ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	# of Analyst Site Visits		*Analyst Forecast Error*	
	(1)	(2)	(3)	(4)
HSR*Soft information	0.033**		-0.025*	
	(0.016)		(0.013)	
HSR* High return volatility		0.024		-0.034*
		(0.026)		(0.018)
HSR	0.036**	0.037*	-0.005	0.003
	(0.017)	(0.006)	(0.012)	(0.006)
Control variables	Yes	Yes	Yes	Yes
Broker-firm pair FE	Yes	Yes	Yes	Yes
Firm city × Year FE	Yes	Yes	Yes	Yes
Brokerage city × Year FE	Yes	Yes	Yes	Yes
Observations	26,146	28,361	135,767	145,112
R-squared	0.569	0.553	0.633	0.611

Table 8. Survey of Remote and Mega-City Analysts

Mega-city analysts are those employed by a broker located in Beijing, Shanghai, Shenzhen, and Guangzhou. Others are classified as remote respondents. Geographical distribution is in Panel A. Panel B summarizes the responses to each preliminary question. Panels C and D summarize the responses to general and specific questions by remote respondents and mega-city respondents, respectively. For all questions in Panels C and D, a 5-point scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*). We reverse the order for the 50 responses that use the alternative survey questions among remote analysts. The % Agree column represents the percent of respondents that agreed or strongly agreed with the statement. A t-test is used to determine if the mean response is statistically different from the neutral midpoint of 3. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Geographical Distribution of Remote and Mega-City Analysts

	Respondents	
City	Number	Percent
Mega-City Respondents (N=236)		
Beijing	118	50.00%
Shanghai	78	33.05%
Shenzhen, Guangdong	30	12.71%
Guangzhou, Guangdong	10	4.24%
Remote Respondents (N=98)		
Wuhan, Hubei	53	54.08%
Chengdu, Sichuan	20	20.41%
Changsha, Hunan	13	13.27%
Qingdao, Shandong	3	3.06%
Jiaxing, Zhejiang	3	3.06%
Jiangyin, Jiangsu	3	3.06%
Xiamen, Fujian	1	1.02%
Harbin, Heilongjiang	1	1.02%
Jiangmen, Guangdong	1	1.02%

Table 8. Survey Evidence (Continued)

Panel B: The Frequency of Remote and Mega-City Analyst Responses to Analyst Profile Questions (N=334)

		Yes		No	
Number of firms following <=8 (Yes/No)	remote	91 (92.86%)		7 (7.14%)	
	mega-city	126 (53.39%)		110 (46.61%)	
		Yes		No	
Direct HSR connection between brokerage and listed firm (Yes/No)	remote	78 (79.59%)		20 (20.41%)	
	mega-city	170 (72.03%)		66 (27.97%)	
Visit local firms more than nonlocal firms (Yes/No)		Yes		No	
	remote	83 (84.69%)		15 (15.31%)	
	mega-city	160 (67.8%)		76 (32.2%)	
		#visits =0	1-3	3-5	#visits =>5
Number of visits per year paid to portfolio firms	remote	6 (6.12%)	55 (56.12%)	29 (29.59%)	8 (8.16%)
	mega-city	12 (5.08%)	121 (51.27%)	60 (25.42%)	43 (18.22%)
		Manufacturing	High-tech	Other industries	
Industry distribution of portfolio firms (Manufacturing, High-tech, and Other industries)	remote	48 (48.98%)	34 (34.69%)	16 (16.33%)	
	mega-city	101 (42.8%)	57 (24.15%)	78 (33.05%)	

Table 8. Survey Evidence (Continued)

Panel C: Responses from Remote Respondents (N=98)

General questions:			
Direct HSR connection results in analysts…	% Agree	Mean	St. Dev.
Conducting more site visits in person	95.92%	4.59***	0.61
Obtaining more firm-specific information	92.86%	4.47***	0.72
Better understanding strategies, operation, and performance	88.78%	4.37***	0.8
Better understanding key challenges/issues facing companies	86.73%	4.34***	0.9
Better understanding corporate culture and employees' morale	87.76%	4.35***	0.9
Specific questions:			
The introduction of a direct HSR route will…			
Increase your frequency of visit	91.84%	4.41***	0.87
Increase your flexibility to visit when most useful	95.92%	4.52***	0.76
Help you better understand current strategies, operation, performance, corporate culture, and employees' morale	86.73%	4.32***	0.84
Enable you to talk to non-management employees to gain more insights into the company's strategies, operations, culture, and employment morale	88.78%	4.37***	0.83
Help you make more accurate earnings forecasts	81.63%	4.18***	0.99
Help you forecast the company's long-term growth	77.55%	4.03***	1.16

Table 8. Survey Evidence (Continued)

Panel D: Responses from Mega-City Respondents (N=236)

General questions: **Direct HSR connection results in analysts…**	% Agree	Mean	St. Dev.
Conducting more site visits in person	86.86%	4.13***	0.79
Obtaining more firm-specific information	77.54%	3.93***	0.84
Better understanding strategies, operation, and performance	67.80%	3.77***	0.92
Better understanding key challenges/issues facing companies	65.25%	3.71***	0.92
Better understanding corporate culture and employees' morale	74.15%	3.86***	0.84
Specific questions:			
The introduction of a direct HSR route will…			
Increase frequency of visit	83.47%	4.10***	0.8
Increase flexibility to visit when most useful	89.83%	4.19***	0.76
Better understand current strategies, operation, performance, corporate culture, and employees' morale	66.10%	3.77***	0.9
Talk to non-management employees to gain more insights into the company's strategies, operations, culture, and employment morale	73.31%	3.88***	0.87
Make more accurate earnings forecasts	48.73%	3.44***	1.04
Help forecast the company's long-term growth	46.19%	3.37***	1.09

Table 9. HSR Introduction and Analyst Coverage Initiation Decision

This table reports the results of testing the effect of HSR introduction on analyst coverage initiation decision. We start with all firms that are included in the CSMAR database. For a broker-firm-year to be included in the coverage initiation sample we require the broker issues no earnings forecast in the prior three years from t-3 to t-1. The dependent variable, *Initiation*, is a dummy variable, equal to 1 if a broker issues a forecast for firm *i* in year *t*, and 0 otherwise. *HSR* is an indicator variable that takes the value of 1 if there is at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. Control variables include firm size, leverage, ROA, book-to-market, return volatility. All variables are defined in Appendix A. We include broker-firm-pair fixed effects, firm city-by-year fixed effects, and broker city-by-year fixed effects. Standard errors clustered at the firm level are reported in parentheses. The superscripts ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Initiation	
	All Firms	SZSE firm after 2009
	(1)	(2)
HSR	0.006***	0.006***
	(0.001)	(0.002)
Control variables	Yes	Yes
Broker-firm pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	549,607	281,238
R-squared	0.337	0.362

Table 10. Cross-sectional variation of the HSR effect on forecast errors in corporate governance

This table examines whether the effect of HSR introduction between a firm-brokerage pair on analyst forecast error varies with firm corporate governance. Two indicator variables serve as the proxy for corporate governance. *SOE* takes the value of one if the firm is a state-owned enterprise in the year immediately before the HSR introduction, and 0 otherwise. *Low board independence* takes the value of 1 if the proportion of independent board of director is below the sample median in the year immediately before the HSR introduction, and 0 otherwise. The dependent variable, *Analyst Forecast Error*, is the natural logarithm of one plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. *HSR* is an indicator variable that takes the value of 1 if there is at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. Firm-broker pair fixed effects and city-times-year fixed effects are included as controls. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Analyst Forecast error	
	(1)	(2)
HSR*SOE	-0.056**	
	(0.026)	
HSR*Low board independence		-0.083**
		(0.038)
HSR	0.013	-0.006
	(0.017)	(0.010)
Control variables	Yes	Yes
Pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	141,104	145,418
R-squared	0.614	0.611

Figure 1. National Development of HSR Network in China

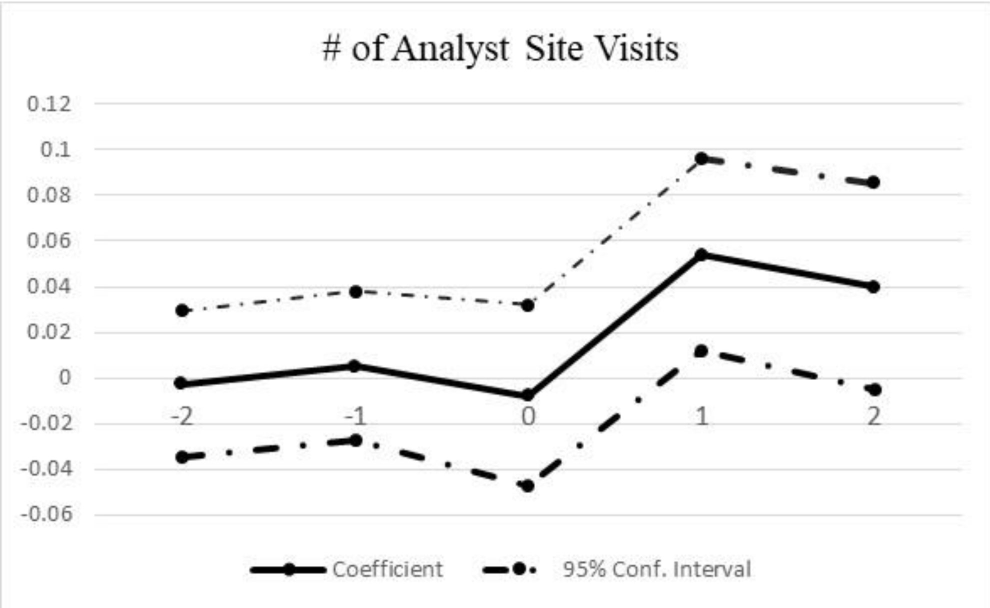
The table and figure depict the introduction and development of high-speed rails (HSR) in China.

Time Period	HSR Lines	Operating mileage(km)
Before August 2008	Conventional railway network, no HSR service	0
Between August 2008 and July 2011	Preliminary construction	6963
Between July 2011 and January 2013	Network skeleton: Several new HSR lines were built	12344
Between January 2013 and December 2018	Deep intensification	29000

Figure 2. Testing the Parallel-Trend Assumption

Panel A: Testing the Parallel-Trend Assumption for Analyst Site-Visits Analysis

Panel A reports the point estimates for *HSR* from an OLS regression. The dependent variable is *# of Analyst Site Visits* in logs. The model specification is the same as column (1) of Appendix C. Point estimates are reported for two years before, one year before, the year of, one year after, and two or more years after the introduction of HSR between a firm and her broker. Ninety-five percent confidence intervals, adjusted for clustering at the firm level, are also plotted.



Panel B: Testing the Parallel-Trend Assumption for Forecast-Errors Analysis

Panel B reports the point estimates for *HSR* from an OLS regression. The dependent variable is *Analyst Forecast Error* in logs. The model specification is the same as column (2) of Appendix C. Point estimates are reported for two years before, one year before, the year of, one year after, and two or more years after the introduction of HSR between a firm and her broker. Ninety-five percent confidence intervals, adjusted for clustering at the firm level, are also plotted.

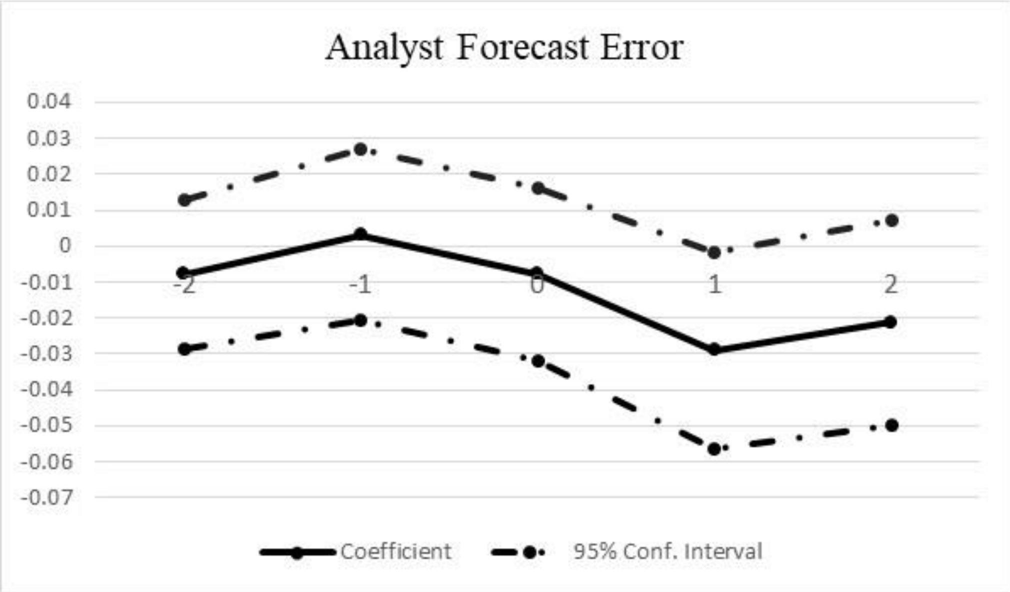


Figure 3 Survey Responses

Panel A: Survey Responses among Remote Respondents

This figure shows the distribution of combined responses to the original and alternative survey questions provided in Internet Appendix IA.5 and IA.6, respectively, among remote analysts. Remote analysts are those employed by a broker located in cities other than Beijing, Shanghai, Shenzhen, and Guangzhou. The first five rows show the responses to the general questions, and the last six rows for the specific questions. On the horizontal axis, positive (negative) percentages refer to "strongly agree," "agree," and "neutral" ("disagree" and "strongly disagree") responses.

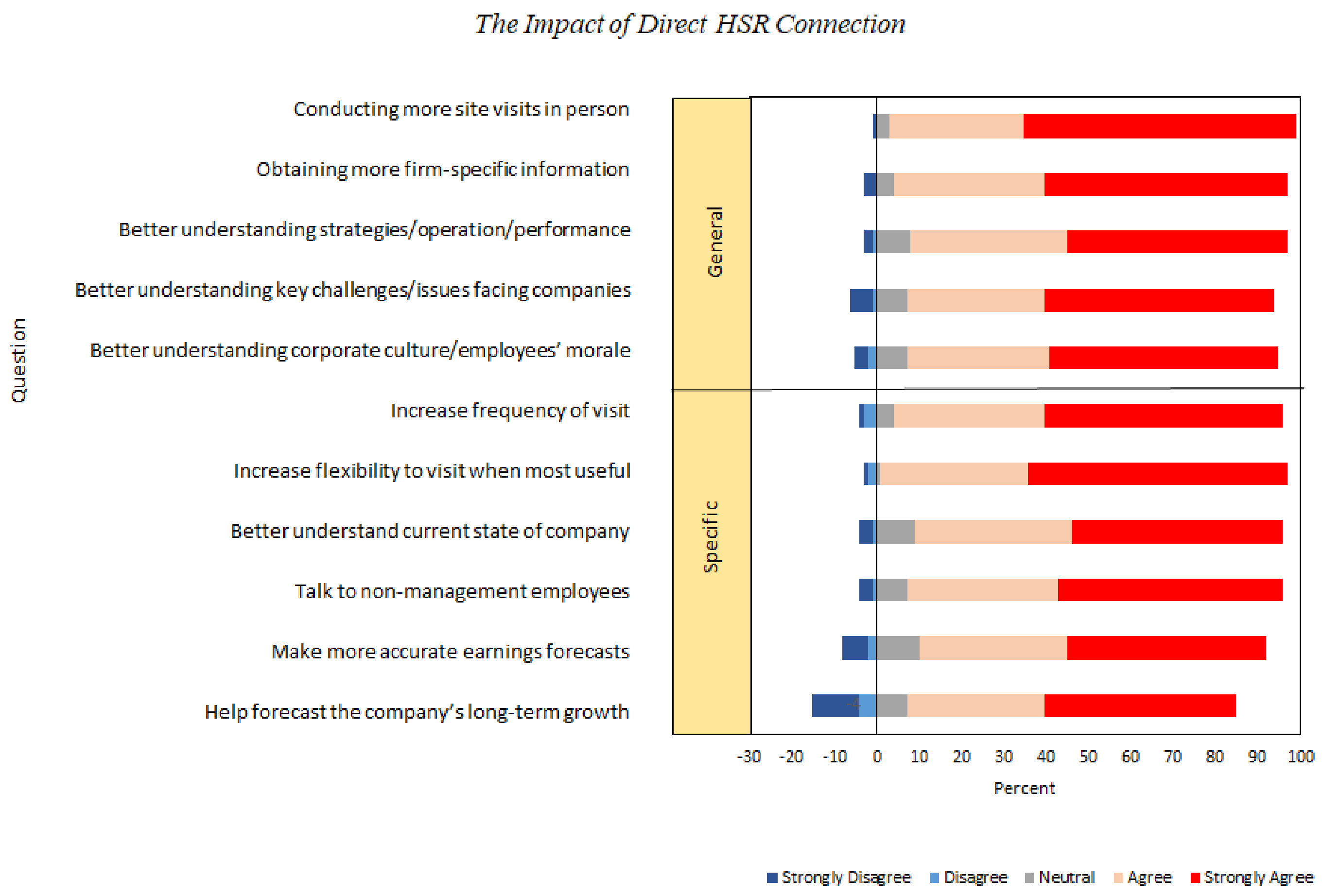
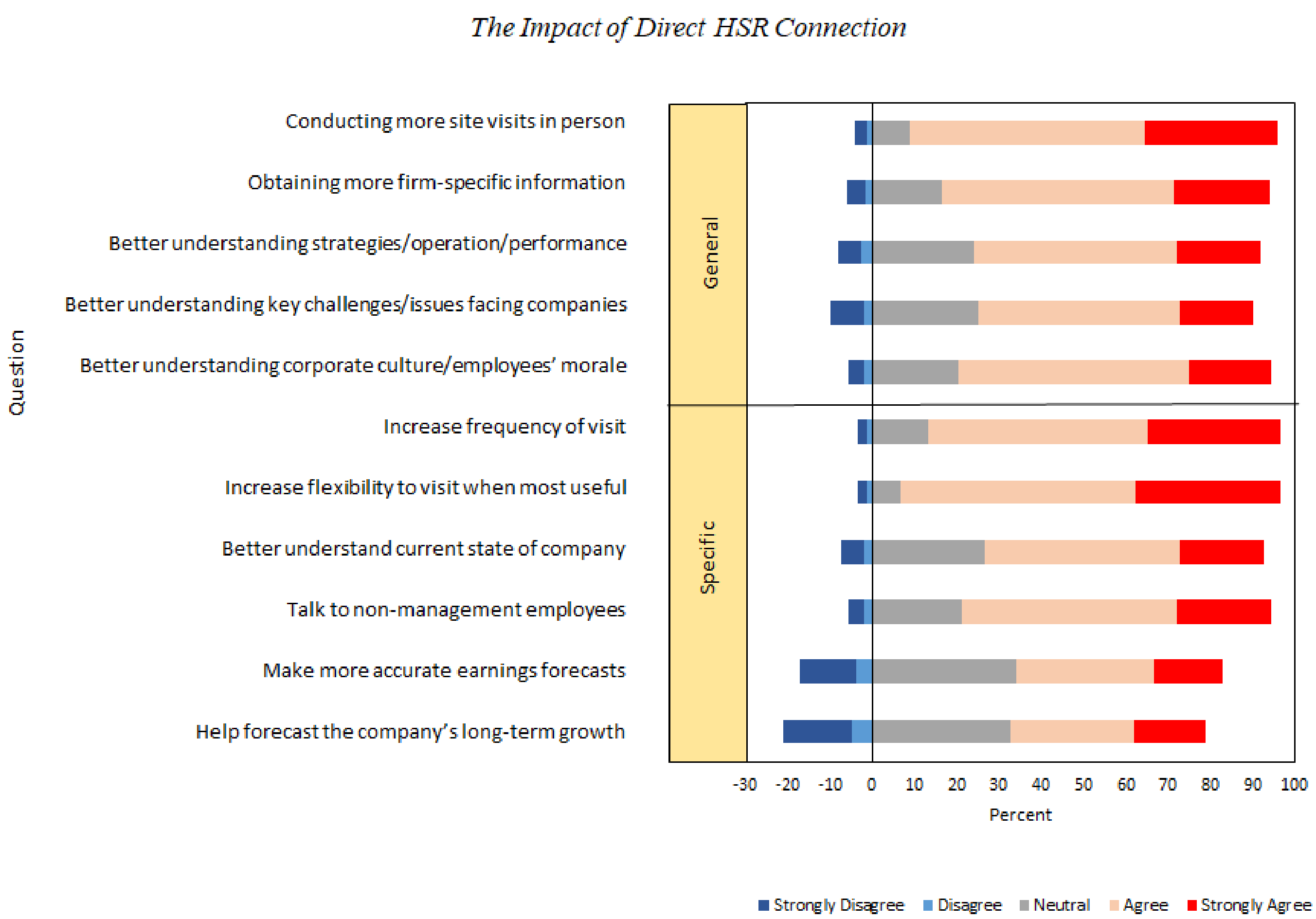


Figure 3 Survey Responses (continued)

Panel B: Survey Responses among Mega-City Respondents

This figure shows the distribution of responses to the original survey questions provided in Internet Appendix IA.5 among mega-city analysts. Mega-city analysts are those employed by a broker located in Beijing, Shanghai, Shenzhen, and Guangzhou. The first five rows show the responses to the general questions, and the last six rows show the responses to the specific questions. On the horizontal axis, positive (negative) percentages refer to "strongly agree," "agree," and "neutral" ("disagree" and "strongly disagree") responses.



Appendix A. Variable Definitions

This table defines all variables used in the empirical analyses.

Variable	Definition
Analyst Forecast Error	The absolute difference between the EPS forecast and actual EPS divided by the share price at the beginning of the year and multiplied by 100 (in logs).
Book-to-market	Book value of equity divided by market value of equity.
Initiation	An indicator variable that equals 1 if a broker issues a forecast for firm *i* in year *t* but issues no forecast in t-3 to t-1, and 0 otherwise.
Difficult to visit	An indicator variable that equals 1 if the distance between the listed firm and the brokerage house is longer than 800 km and if no direct flight exists between the firm and the analyst by 2005, and 0 otherwise.
EAINFO	Size-adjusted absolute cumulative abnormal returns for the three-day window around the annual earnings announcement date and multiplied by 100 (in logs).
Firm experience	The average firm-specific experience of all analysts in a broker team covering the firm (in logs).
Forecast duration	Number of days between forecast issuance date and current earnings-announcement date (in logs). For example, an earnings forecast made on May 1, 2010, and the corresponding actual earnings announced on February 28, 2011, will have a duration of 303 days, and an earnings forecast made on November 1, 2010, will have a duration of 119 days. This variable has a mean of 253.3 days, median of 242 days, and standard deviation of 85.996 days.
FRINFO	We first calculate the size-adjusted cumulative abnormal returns for the three-day window around the forecast revision date. For earnings-forecast revision <0, we multiply raw CAR by -1; for new earnings forecast or earnings revision>=0, we multiply raw CAR by 1 (in logs).
HSR	An indicator variable that takes the value of 1 if at least one HSR connects the firm city and the brokerage city directly at the beginning of year *t*, and 0 otherwise.
Leverage	Book value of debt divided by the book value of total assets.
PHSR1	The proportion of analysts who are connected to the firm by an HSR at the beginning of year t, relative to the total number of analysts covering the firm in t.
PHSR2	The proportion of research reports signed by HSR-connected analysts relative to the total number of research reports signed by all analysts covering the firm in t. An analyst is HSR connected if the analyst is directly connected with the portfolio firm by an HSR at the beginning of t.
Return volatility	The standard deviation of monthly stock return over the year.
ROA	Return on assets, measured as net income over book value of total assets.
Soft information	Within a firm-year, we classify all analysts for that firm-year into distant and near groups based on the median distance between the analyst and the firm. We then calculate the average EPS forecast errors of respective group over t-3 and t-1. Soft information takes the value of 1 if the average EPS forecast errors for the near group is lower than that for the distant group, and 0 otherwise.
Market reaction to stock recommendation	We first calculate the buy-and-hold return over the trading period [-1, +1], subtracted by the corresponding buy-and-hold value-weighted market return for the same period, where 0 represents the recommendation issuance date. For recommendation revisions made within one year, we invest $1 in the firm stock with recommendation upgrades, and short $1 with recommendation downgrades. If no previous recommendation is made in one year or the current recommendation reiterates the prior one longer than a year, we treat the current recommendation as an "initiation." For recommendation initiations or reiterations within a year, we invest $1 for "Strong Buy" and "Buy," $0 for "Neutral," and short $1 for "Strong Sell" or "Sell." Last, we multiple buy-and-hold return with the investment amount to obtain the profitability measure (in logs).
Total assets	Book value of total assets (in logs).

# of Analyst Site Visits	The number of site visits conducted by analysts in brokerage j for firm i in year t (in logs).
# of analyst following	The total number of brokers covering a firm in year t (in logs).

Appendix B. Full Set of Results for Site-Visits and Forecast-Errors Analyses

This table reports the full set of results of OLS regressions that examine the effect of HSR introduction between a firm-brokerage pair on analyst site visits and their earnings-forecast quality. *HSR* is an indicator variable that takes the value of 1 if at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. In column (1), the dependent variable, *# of Analyst site Visits*, is the natural logarithm of 1 plus the number of site visits conducted by a brokerage for a firm in year t. In column (2), the dependent variable is *Analyst Forecast Error*, which is the natural logarithm of 1 plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. Control variables are defined in Appendix A. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	# of Analyst Site Visits)	*Analyst Forecast Error*
	(1)	(2)
HSR	0.049***	-0.021**
	(0.014)	(0.009)
Total assets	0.028	-0.010
	(0.031)	(0.024)
Leverage	-0.158**	-0.028
	(0.080)	(0.077)
ROA	-0.412**	-2.432***
	(0.161)	(0.299)
Book-to-market	-0.112***	0.150***
	(0.040)	(0.046)
Return volatility	0.148	0.616***
	(0.104)	(0.149)
# of analyst following	0.109***	-0.053***
	(0.014)	(0.016)
Firm experience	-0.029***	-0.003
	(0.010)	(0.006)
Forecast duration		0.415***
		(0.009)
Broker-firm Pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	31,417	160,519
R-squared	0.551	0.612

Appendix C. Testing the Parallel-Trend Assumption

This table examines pre-treatment trends between the treated group and the control group for site-visits and forecast-error analyses. In columns (1) and (2), the regression specification is the same as column (1) of Table 2 and Table 3, respectively, except we replace the indicator *HSR* with the indicators HSR^{-2}, HSR^{-1}, HSR^{0}, HSR^{+1}, and HSR^{2+}. These five indicators flag the year relative to the introduction of HSR between firm i and brokerage j in year 0. In particular, HSR^{-2}, HSR^{-1}, HSR^{0}, HSR^{+1}, and HSR^{2} indicate, respectively, two years before, one year before, the year of, one year after, and two or more years after the HSR introduction. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	# of Analyst Site Visits	*Analyst Forecast Error*
	(1)	(2)
HSR^{-2}	-0.002	-0.008
	(0.016)	(0.011)
HSR^{-1}	0.006	0.003
	(0.017)	(0.012)
HSR^{0}	-0.007	-0.008
	(0.020)	(0.012)
HSR^{+1}	**0.054***	**-0.029***
	(0.021)	**(0.014)**
HSR^{2+}	**0.040***	**-0.021**
	(0.023)	**(0.015)**
Control variables	Yes	Yes
Pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	31,417	160,519
R-squared	0.551	0.612

Appendix D. The Payoff of HSR Introduction to Brokers and Analysts

We find that HSR introduction significantly increases analyst site visits and improves their forecast accuracy. The economic magnitude is also significant: a 4.9% increase in the number of annual site visits and 2.1% increase in forecast accuracy. This result indeed raises the question why analysts did not undertake these site visits before the introduction of HSR, if the reduction in information acquisition costs leads to large payoff to both brokers and analysts.

A thorough answer to this question requires complete information about the costs and benefits of a visit at any point in time, (including for example the shadow price of analysts' time). This information is not available to us or to other researchers. While we are unable to give a definitive answer to this question, we take several significant strides addressing this issue—even within the framework of the HSR introduction.

As said, ideally, to assess the payoff of site visits we would like to observe the profit function of both the brokerage firms and analysts over time; and then evaluate how the brokerages' profits and the analysts' personal compensation have been affected by the introduction of HSR. While a complete data is not available, we use several proxies for the increased profitability of site visits. Regarding brokerages we examine the increased trading volume associated with the HSR introduction. Using the estimated increase in trading volume we also calculate the increase in trading commissions. Since the incentive to optimize the number of visits is not only from the broker perspective but also from the analysts, we also examine two available aspects related to analysts' compensation: demotions and promotions (data on analysts' direct compensation is not available).

The payoffs of HSR introduction to brokers

Our objective here is to assess the increase in commission revenues associated with HSR introduction. To this end, we first estimate the increased turnover around the announcement of analyst forecast revisions. Specifically, we measure stock turnover during a two-day window [0, +1]. Turnover is calculated as RMB turnover in those two days relative to market value and expressed in percentage. Stock turnover data are obtained from the China Stock market & Accounting Research (CSMAR) database that covers Chinese publicly traded firms listed on Shanghai (SHSE) and Shenzhen (SZSE) stock exchanges. We then merge the stock turnover data with the forecast revision sample. Recall we exclude revisions issued on the same day and the next day when earnings are announced. Our sample holds 53,409 forecast revisions over the period of 2005-2019. Panel A shows the mean two-day turnover around forecast revisions is 4.8% (equivalent to the 2.4% daily turnover), which is statistically and economically higher than the average daily turnover of 2.3% during the non-event window, suggesting that forecast revisions have information content.

To examine the payoffs of HSR introduction to brokers, we again use difference-in-differences approach based on the following equation.

$$y_{a,i,t} = \beta_0 + \beta_1 HSRDC_{a,i,t} + Controls_t + \alpha_{a,i} + \alpha_{city(i)} \times \alpha_t + \alpha_{city(a)} \times \alpha_t + \varepsilon_{a,i,t},$$

where *a* indexes brokerages, *i* indexes firms, *t* indexes years, *city(i)* and *city(a)* index the city where firm *i* and brokerage *a* are located, respectively, and $y_{a,i,t}$ is stock turnover around the announcement of forecast revisions (in logs) in year t. $HSR_{a,i,t}$ takes the value of 1 for year t if at least one HSR connects firm *i*'s city and broker *a*'s city at the beginning of t. Controls include firm size, leverage, ROA, book-to-market, return volatility, the number of analysts following, and an analyst's firm-specific experience. These control variables are the factors identified in prior studies to affect stock turnover (Chae 2005; and Lo and Wang 2006). We use the same fixed effects structure as for the forecast error analysis. The main coefficient of interest, β_1, captures the change in stock turnovers around the announcement of forecast revisions in response to HSR introduction between the broker's and firm's city. If HSR introduction reduces acquisition costs, which affects the amount of information collected and information quality, and in turn the payoff to brokers, we expect β_1 to be positive. To allow for serial dependence of the error terms, we cluster standard errors at the firm level.

Panel B shows a positive and statistically significant coefficient on *HSR* after controlling for variables associated with stock turnover.[20] To assess the RMB profit to the broker due to HSR introduction, we use the change in turnover around the forecast release. The mean market value of firms in our sample is RMB16.425 billion and the average number of forecast revisions by an analyst for a firm year is 3. Based on prior studies we use a commission fee of 0.25% (Chen and Jiang 2017). With these numbers, and assuming all the abnormal volume associated with the research report release flows to the broker who issues it, the HSR introduction results in RMB 260,172 ($37,164) increase in annual commission fee.[21] If only half the volume flow through the issuing broker, then the fees naturally are cut by half. Overall, the gains seem meaningful, and are at the range of about 25% of the annual compensation of an analyst with at least four years of experience and ranked the top 10% in China;[22] but they are not at the order of magnitude that suggest brokerage houses left significant gains on the table prior to the HSR.

The payoffs of HSR introduction to analysts

Ideally, we would have been able to obtain analysts' total compensation both before and after the HSR introduction. Given that this information is unavailable, we instead assess the payoff to analysts due to HSR introduction based on the commission they received from their employers. Our conversation with Chinese brokers indicates that part of analysts' compensation comes from trading commission (roughly 20% of trading commission that their employers receive). Based on this number together with the results of stock turnover analysis and that on average 1.7 analysts from each broker cover a firm, HSR introduction increases analysts' annual payoff by RMB 30,608 ($4,372). Given analysts' annual compensation of roughly RMB 1 million ($142,000), this benefit of about 3% is meaningful but not excessively large.

Besides commission fee, HSR introduction might benefit analysts in their career outcomes. Following Hong, Kubik and Solomon (2000) and Hong and Kubik (2003), we analyze the incremental analysts' payoff from HSR introduction through its outcome on two aspects of their career: promotions and demotions. These outcomes are likely to be correlated with their career objective and compensation. If an analyst moves to a brokerage house that is more prestigious than the broker she is currently employed, we consider it a job promotion. We use broker size measured by the number of analysts employed as a proxy for its prestige (e.g. Michaely and Shaw 1994). Based on broker size, we classify brokerage houses in our sample into two groups each year: prestigious brokerage houses which rank among top 10, and less prestigious brokerage houses which are outside the top 10 group. While the choice of top 10 might be arbitrary, our objective is to separate prestigious brokers from the less prestigious ones so that we can capture analysts' career movement more precisely.[23] We then code a dummy variable, *promotion*, equal to 1 if an analyst moves from a low status broker to a more

[20] We also conduct the analysis using 2SLS estimation based on equations (1) and (2), where stock turnover around forecast revision announcements serves as the dependent variable. The coefficient estimate on *HSR* in the first stage is almost identical to that reported in column (1) of Table 6 (Coefficient =0.056), suggesting that HSR introduction increases site visits. The coefficient on the predicted analyst site visits in the second-stage regression is 1.563, and statistically significant at the 5% level, suggesting that site visits increase stock turnover around analyst forecast revisions. This inference is consistent with that from the reduced form estimation, but again uses a limited sample.

[21] Annual commission is computed as 4.8%*4.4%*16.425 billion *0.25%*3= RMB 260,172.

[22] Analyst Similar to Institutional Investor's "All American Research team" rankings, New Fortune provides ranking for Chinese analysts. According to New Fortune (2019), the annual compensation for analysts who entered the final list (7-8 analysts per industry for roughly 30 industry sectors) is roughly RMB1 million and the top ranked analysts earned RMB3-4 million.

[23] It is important to note all results are insensitive to whether we code demotion and promotion based on the quartile ranks of broker size or raw broker size.

prestigious broker in year t, and 0 otherwise. Conversely, *demotion*, takes a value of 1 if an analyst moves from a prestigious broker to a low status broker, or out of the profession in year t, and 0 otherwise.

Hong et al. (2000) find that forecast accuracy affects promotion only among analysts whose forecast performance is poor. They also find that analysts from small brokers have lower forecast performance than their counterparts from large brokers. These finding imply that the effect of forecast accuracy due to HSR introduction on promotion is likely to be stronger for analysts from small brokers. Therefore, we conduct our analyses for small and large brokers separately. We classify a broker above (below) the sample median size as large (small). The summary statistics reported in Panel C show the annual percentage of analysts being promoted and demoted is 2.3% and 15.2% for small brokers, and 1.0% and 16.8% for large brokers, slightly lower than the numbers (3.29% for promotion and 17.17% for demotion) reported in Table 2 of Hong et al. (2000) based on the U.S. analysts.

Following Hong and Kubik (2003), we analyze the payoffs of HSR introduction to analysts at the analyst-year level and estimate the following model:

$$\begin{aligned}\Pr\left(Promotion\ or\ Demotion_{a,t+1} = 1\right) \\ = \beta_0 + \beta_1 PHSR_{a,t} + \beta_2 PHSR_{a,t} * \text{Small broker} + \alpha_i + \delta_t + \varepsilon_{a,t},\end{aligned}$$

where P*romotion* and *Demotion*, as defined above, serve as the dependent variable, alternatively. $PHSR_{a,t}$ measures the intensity of HSR connections easing analyst *a*'s commuting to her portfolio firms. We construct this measure based on the number of firms she covers that are HSR connected to her at the beginning of year t.[24] For example, if an analyst covers 10 firms in year t and three of them are connected to her by HSR at the beginning of t, then *PHSR* is equal to 0.3 (=3/10). The higher the value, the easier for the analyst to travel to her portfolio firms. α_i are broker fixed effects, and δ_t are year fixed effects. We estimate the equation using linear probability model with standard errors clustered at the analyst level. If HSR eases analysts' travel to portfolio firms and improves forecast accuracy, we expect the intensity of HSR connection to portfolio firms is positively associated with the promotion likelihood and negatively associated with the demotion likelihood.

Panel D presents the results. Column (1) focuses on the promotion likelihood. We find a positive and statistically insignificant coefficient on *PHSR*, suggesting that the intensity of HSR connections to analysts' portfolio firms has little effect on their promotion likelihood among analysts from large brokers. However, we find a significant HSR effect on promotion for analysts employed by small brokers as evident by the positive and significant coefficient on the interaction term (*PHSR*Small broker*). This result is consistent with Hong et al. (2000)'s findings that forecast performance matters for promotion only among analysts whose forecast accuracy is at the bottom of distribution. For an analyst employed by a small broker, given that the standard deviation of *PHSR* is 0.351, a one standard deviation increase in this variable improves her promotion probability by 0.56% (=0.351*(0.014+0.002)). Considering the 2.3% mean promotion likelihood, this number is economically significant.

When we focus on demotion, we find significant HSR effect for analysts from both large and small brokers. Column (2) shows a negative and statistically significant coefficient on *PHSR*, suggesting that HSR connection intensity reduces analysts' demotion likelihood among large brokers. A one standard deviation increase in *PHSR* reduces the demotion probability by 1.6% (=0.351*0.046). We further find a negative and significant coefficient on *PHSR*Small broker*, implying that the benefit of HSR connections is even larger among analysts from small brokers. A one standard deviation increase in *PHSR* reduces their demotion probability by 4.2% (=0.351*(0.046+0.074)).

[24] Our results are robust to using an alternative measure based on analyst research output. That is, the number of research reports issued to the HSR-connected portfolio firms relative to the number of research reports issued to all portfolio firms by that analyst.

In sum, our evidence suggests that HSR connections benefit analysts by increasing the compensation they earn from commission fees. We also find that the intensity of HSR connections to their portfolio firms significantly reduces unfavorable career outcomes across both large and small brokers, and improves their favorable career outcome for analysts from small brokers.[25]

Finally, one might also consider the dynamics of analysts' labor supply. During the time of our study, the number of analysts in China was in shortage. For example, Wu (2014) notes that about 10,000 chartered financial analysts are needed nationwide over the period of 2010-2017. However, based on the statistics from the CFA Asia Pacific Association, only 1,800 charter analysts were available as of the end of 2010. This shortage suggests that skilled analysts are likely a scarce resource, and thus difficult for brokerages to hire additional analysts to conduct site visits to harvest the benefits. Furthermore, the shortage of analysts' labor supply can also be seen by the large portfolio of firms covered by an analyst. For example, in the U.S. the broker to firm ratio is 0.05-0.06 over our sample period (Merkley et al., 2017), which is much higher than their Chinese counterpart − 0.037 (=86/2300) (Report of Chinese brokerage and securities industry 2012). This lower broker-to-firm ratio suggests that it might be difficult for analysts to conduct site visits without HSR, particularly among remote firms for which site visits are time consuming and among time-constrained analysts (i.e., experienced analysts).

The effect of HSR introduction on stock turnover around analyst forecast revisions

Panel A reports summary statistics of cumulative stock turnover and daily turnover over a two-day window [0,+1] around forecast revisions (TO), and Panel B examines the effect of HSR introduction on the stock turnover (TO). *HSR* is an indicator variable that takes the value of 1 if at least one HSR connects the firm-broker pair at the beginning of *t*, and 0 otherwise. The dependent variable, *TO*, is cumulative value traded in RMB over a two-day window [0,+1] around forecast revisions, scaled by market value (in logs). Control variables include total assets (in logs), leverage, ROA, book-to-market, return volatility, number of analysts following (in logs), analyst firm experience (in logs), and forecast duration. Control variables are defined in Appendix A. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A Summary statistics of stock turnover

Variable	# of Obs.	Mean	Std. Dev.	Median
TO [0,1]	53409	0.048	0.053	0.031

[25] To directly examine the career impact of site visits on analysts, we also conducted 2SLS estimation, where we instrument analyst site visits by *PHSR* in the first stage, and regress analyst promotion or demotion indicator on the predicted site visits obtained from the first-stage regression along with other control variables. The untabulated results show a positive and statistically significant coefficient on *PHSR* (coefficient = 0.223) in the first stage, suggesting that a one standard deviation increase in *PHSR* results in 7.8% increase in site visits (=0.223*0.351*100%, where the standard deviation of *PHSR* is 0.351 as shown in Appendix D Panel C). The coefficient on the predicted site visits in the second stage is positive when promotion serves as the dependent variable (coefficient = 0.066) and negative when demotion serves as the dependent variable (coefficient =-0.392). Both are statistically significant. These results are consistent with those obtained from the reduced form estimation, suggesting that site visits are favorable to analysts' career outcome by improving their promotion likelihood and decreasing demotion likelihood.

Panel B: HSR introduction and stock turnover

	TO
HSR	0.044*
	(0.026)
Control variables	Yes
Pair FE	Yes
Firm city × Year FE	Yes
Brokerage city × Year FE	Yes
Observations	47,988
R-squared	0.605

The effect of HSR introduction on analyst career outcomes

Panel C reports summary statistics of analyst job movement. *Promotion* is an indicator variable, equal to 1 if an analyst moves from a low status brokerage house to a more prestigious brokerage house, and 0 otherwise. Brokers are ranked annually. If a broker ranks in top 10 based on the number of analysts employed in t, it is considered as a prestigious broker, and other brokers are considered as low status. *Demotion* is coded as 1 if an analyst moves from a more prestigious house to a low status one, or moves out of the profession in t or t+1. Panel D examines the effect of HSR connection on analysts' career outcomes at the analyst-year level. *PHSR* is the proportion of firms covered by an analyst that she is connected to by an HSR at the beginning of t, relative to the total number of firms covered by her in year t. t. *Small broker* is an indicator variable that equals 1 if the broker's ranking is below the sample median, and 0 otherwise. Standard errors clustered at the analyst level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel C Summary statistics

Variable	# of Obs.	Mean	Std. Dev.	Median
PHSR	15915	0.341	0.351	0.250

Variable	Small brokers (N=7545)		Large brokers (N=8370)	
	Mean	Std. Dev.	Mean	Std. Dev.
Promotion	0.023	0.151	0.010	0.101
Demotion	0.152	0.359	0.168	0.374

Panel D: HSR introduction and Analyst career outcomes

	Promotion	*Demotion*
	(1)	(2)
PHSR	0.002	-0.046***
	(0.005)	(0.016)
PHSR*Small Broker	0.014**	-0.074***
	(0.006)	(0.018)
Small Broker	0.001	0.003
	(0.004)	(0.011)
Control variables	No	No
Broker FE	Yes	Yes
Year FE	Yes	Yes
Observations	15,913	15,913
R-squared	0.020	0.025

Appendix E. Validation and Robustness Test of the Soft-Information Measure

Though theoretically grounded and intuitive, one might still be concerned about the validity of the soft-information measure. One concern is whether the difference in the average distance to the covered firm between near and distant analysts (hereafter differential distance) are economically large enough to produce differential forecast accuracy. Untabulated results show the median differential distance is 885 km with the 25th percentile of 637 km and 75th percentile of 1,114 km, suggesting the differential distance between near and distant analysts is indeed economically meaningful enough for travel time (and cost) to influence their information acquisition.

The second concern is whether our measure simply reflects the impact of distance, rather than the differential distance between the two broker groups, on forecast accuracy. For example, all analysts who cover a firm might be fairly close to the firm, compared to another firm whose analysts are all far away. Results (untabulated) show the median differential distance is 879 km for firms whose soft information is important and 890 km for firms whose soft information is not important. Thus, our soft-information measure is unlikely to simply reflect the distance between the firm and the average broker covering the firm. In conclusion, our analysis suggests the soft-information measure captures the importance of soft information in valuing a firm.

To further validate our soft-information measure, we correlate it with one firm characteristics likely to be associated with firms' soft information: stock return volatility. Firms with higher stock return volatility commonly have more growth opportunities. Because growth opportunities are inherently more difficult to value than assets in place, face-to-face interaction with top managers and employees are thus important for analysts to assess project quality, and thus soft information becomes important when firms higher stock return volatility. Indeed, we find the soft-information measure is positively significantly associated with monthly stock return volatility (correlation coefficient = 0.008, p-value = 0.004), suggesting our soft-information measure likely captures the underlying importance of soft information.

The Effect of HSR Introduction and the Importance of Soft Information: An Alternative Measure of Soft Information

This table examines the heterogeneous effect of HSR introduction on analyst site visits and forecast errors based on an alternative measure of soft information. *HSR* is an indicator variable that takes the value of 1 if there is at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. The *soft-information category* is a categorical variable, equal to the quartile ranked difference in the average EPS forecast errors between the near and distant analyst group for a given firm, with higher values representing forecasts issued by near analysts being more accurate than those issued by distant analysts. That is, it takes values of 3, 2, 1, and 0 if the difference in the average EPS forecast errors between the near and distant group is in the bottom 3rd, 2nd, and top quartile, respectively. In column (1), the dependent variable, *# of Analyst Site Visits*, is the natural logarithm of 1 plus the number of site visits conducted by a brokerage for a firm in year t. In column (2), the dependent variable, *Analyst Forecast Error*, is the natural logarithm of 1 plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	# of Analyst Site Visits	*Analyst Forecast error*
	(1)	(2)
HSR*Soft information category	0.016**	-0.021***
	(0.008)	(0.007)
HSR	0.029	0.013
	(0.019)	(0.015)
Soft information category	-0.001	0.017**
	(0.006)	(0.007)
Control variables	Yes	Yes
Broker-firm pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	26,146	135,767
R-squared	0.569	0.633

Appendix F. Distance and Analyst Forecast Errors and Price Efficiency

This appendix presents the results of the relation between the difficulty for analysts to visit a firm and the firm's consensus forecast errors and market reaction to earnings announcements. The dependent variable is the absolute difference between the consensus EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100 in column (1), and market reaction to earnings announcements in column (2), calculated as size-adjusted absolute cumulative abnormal returns for the three-day window (-1, +1) around the annual earnings announcement date and multiplied by 100 Both variables are in logs. The independent variable is *Average difficult to visit*, which is the mean of the difficult to visit across all analysts within a firm. *Difficult to visit* is an indicator variable that equals 1 if the distance between the listed firm and the brokerage house is more than 800 km and no direct flight exists between the firm and the analyst by 2005, and 0 otherwise. Controls include firm size, leverage, ROA, book-to-market, return volatility, and the number of analysts following. All variables are defined in Appendix A. We include year fixed effects. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Consensus Error	*EAINFO*
	(1)	(2)
Average difficult to visit	0.028*	0.029*
	(0.015)	(0.017)
Control variables	Yes	Yes
Year FE	Yes	Yes
Observations	11,549	11,549
R-squared	0.213	0.040

Appendix G. Cross-sectional Heterogeneity in the HSR Effect

In this appendix, we report the results of testing whether the HSR effect on analyst site visits varies with analyst time constraints and prior forecast performance. The results for analyst time constraints are discussed in section 6.3, thus our discussion focuses on analysts' prior forecast performance. Underperforming analysts might have a stronger incentive to take advantage of HSR introduction by visiting portfolio firms more often, in order to improve their forecast performance and in turn career prospect (Hong, Kubik and Solomon 2000). Forecast accuracy is measured by an indicator, *Low accuracy rank*, taking values of 1 if the analyst relative forecast accuracy ranking for the treated analysts in the year immediately before HSR introduction is below the sample median, and 0 otherwise. We follow Hong et al. (2000) to calculate analyst relative ranking. As shown in column (3) below, analysts' relative forecast accuracy before HSR introduction has no impact on the HSR effect in the cross section.

This table examines whether the effect of HSR introduction between a firm-brokerage pair on analyst site visits varies with analysts' time constraints and forecast accuracy before HSR introduction. Two indicator variables serve as the proxy for time constrains. One is *High industry experience* taking the value of 1 if the length of industry experience for the treated analysts in the year immediately before HSR introduction is above the sample median, and 0 otherwise. The other is *Travel constraints* taking value of 1 if the sum of the distances (or the total distance) between each of the portfolio firms and the analyst in the year immediately before HSR introduction is above the sample median, and 0 otherwise. The dependent variable *# of Analyst site visits* is the natural logarithm of one plus the number of site visits. *HSR* is an indicator variable that takes the value of 1 if there is at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. Firm-broker pair fixed effects and city-times-year fixed effects are included as controls. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	# of Analyst Site Visits		
	(1)	(2)	(3)
HSR*High industry experience	0.040*		
	(0.024)		
HSR*Travel constraints		0.032	
		(0.023)	
HSR*Low accuracy rank			0.011
			(0.025)
HSR	0.000	0.009	0.022
	(0.022)	(0.021)	(0.021)
Control variables	Yes	Yes	Yes
Pair FE	Yes	Yes	Yes
Firm city × Year FE	Yes	Yes	Yes
Brokerage city × Year FE	Yes	Yes	Yes
Observations	19,965	19,965	19,897
R-squared	0.555	0.555	0.555

Internet Appendix IA. 1 Sample Filter for Analyst Data

In our empirical tests we only require analysts to follow a firm for at least two consecutive years. For the treated analyst-firm pairs, we further require the analyst cover the firm for at least one year before the treatment (i.e., HSR introduction) and one year after the treatment. We do not impose a requirement of all analysts to be present throughout the sample period for two reasons. First, an analyst stays for a little shorter than four years on average in our sample (similar number is observed in Hong et al. (2000)), and 90% analysts stay for less than 8 years, which is much shorter than the 15 years of our sample period (2005-2019).Therefore, such requirement likely induces look-ahead bias, that is, we might only select those successful analysts ex post. If these analysts tend to be time constrained and thus are more likely to take advantage of HSR introduction (as documented in our response to Comment 4), then the estimated HSR effect can be biased upward in absolute magnitude. Furthermore, our selection criteria are in line with Giroud (2013), who require a plant have a minimum of two consecutive years of data to be included in the sample. Second, given the long sample period relative to the average analyst career span (15 years vs. 4 years), the imposition of constant stay would reduce our sample size to a greater extent, thereby lowing our test power to detect the treatment effect.

Internet Appendix IA. 2 Validation of Recommendations Data quality and Summary Statistics of Market Response to Stock Recommendations

Panel A: Validation of recommendations data

To validate the quality of the recommendations data obtained from CSMAR, we cross-checked recommendations obtained from CSMAR with those from I/B/E/S. We find no discrepancy regarding the issuance date and the level of recommendations between the two sources. We further randomly selected 100 original analyst reports downloaded from INVESTODAY (https://www.investoday.com.cn/) and verified the absence of any coding error in stock recommendations from CSMAR. Thus, we are confident about the quality of our recommendation data.

Panel B: Summary statistics of market response to stock recommendations

This table reports the three-day size-adjusted cumulative abnormal return (CAR) centering on the recommendation issuance date (-1, +1) by recommendation categories. ***, **, and * denote means significantly different from 0 at the 1%, 5%, and 10% levels, respectively.

Category	# of Obs.	Mean	S.D.
Reiterations within one year	34685	1.112***	4.665
Reiterations longer than one year	4161	1.01***	4.580
Recommendation initiations (no prior recommendation)	7230	1.257***	5.025
Recommendation upgrades within one year	2206	2.078***	5.424
Recommendation upgrades longer than one year	1951	1.845***	5.091
Recommendation downgrades within one year	1252	-0.321***	4.665
Recommendation downgrades longer than one year	889	0.304***	4.163

Internet Appendix IA. 3 HSR Introduction, Analyst Site Visits, and Analyst Forecast Errors using Ease-to-Travel and Local Analysts Sample

This table examines the effect of HSR introduction between a firm-brokerage pair on analyst site visits and analyst forecast errors, using firms located in the four mega-cities (Beijing, Shanghai, Guangzhou, and Shenzhen), and analysts located in the same city as their portfolio firms. *HSR* is an indicator variable that takes the value of 1 if at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. The dependent variable in column (1), *# of Analyst Site Visits*, is the natural logarithm of 1 plus the number of site visits conducted by a brokerage for a firm in year t. The dependent variable in column (2), *Analyst Forecast Error*, is the natural logarithm of 1 plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	# of Analyst Site Visits	*Analyst Forecast Error*
	(1)	(2)
HSR	-0.025	-0.006
	(0.035)	(0.016)
Control variables	Yes	Yes
Pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	9,266	53,608
R-squared	0.518	0.484

Internet Appendix IA. 4 Lobby Efforts and the Effect of HSR Introduction on Analyst Site Visits and Forecast Errors

This table examines whether lobby efforts explain the effect of HSR introduction between a firm-brokerage pair on analyst site visits and analyst forecast errors, based on the subsample that excludes the firm and broker-pair cities that are located on the two ends of an HSR in Panel A, and the top 5% largest firms in Panel B, respectively. *HSR* is an indicator variable that takes the value of 1 if at least one HSR connects the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. The dependent variable in column (1), *# of Analyst Site Visits*, is the natural logarithm of 1 plus the number of site visits conducted by a brokerage for a firm in year t. The dependent variable in column (2), *Analyst Forecast Error*, is the natural logarithm of 1 plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Exclusion of firm-broker pair cities located on the two ends of an HSR

	# of Analyst Site Visits	*Analyst Forecast Error*
	(1)	(2)
HSR	0.045***	-0.022**
	(0.014)	(0.009)
Control variables	Yes	Yes
Pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	29,735	151,241
R-squared	0.552	0.614

Internet Appendix IA. 4 Lobby Efforts and the Effect of HSR Introduction on Analyst Site Visits and Forecast Errors (Continued)

Panel B: Exclusion of top 5% largest firms

	# of Analyst Site Visits	*Analyst Forecast Error*
	(1)	(2)
HSR	0.051***	-0.017*
	(0.015)	(0.009)
Control variables	Yes	Yes
Pair FE	Yes	Yes
Firm city × Year FE	Yes	Yes
Brokerage city × Year FE	Yes	Yes
Observations	29,016	149,950
R-squared	0.557	0.615

Internet Appendix IA. 5 HSR Introduction and Analyst Forecast Errors Using One-Year-Ahead Forecasts

This table examines the effect of HSR introduction between a firm-brokerage pair on analyst forecast errors using one-year-ahead earnings forecasts. The unit analysis is at the forecast level. *HSR* is an indicator variable that takes the value of 1 if at least one HSR connecting the firm city and the brokerage city at the beginning of *t*, and 0 otherwise. The dependent variable, *Analyst Forecast Error*, is the natural logarithm of 1 plus the absolute difference between the EPS forecast and actual EPS, divided by the share price at the beginning of the year and multiplied by 100. Standard errors clustered at the firm level are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	1+ Analyst Forecast Error
	(1)
HSR	-0.023*
	(0.012)
Control variables	Yes
Pair FE	Yes
Firm city × Year FE	Yes
Brokerage city × Year FE	Yes
Observations	165,668
R-squared	0.662

Internet Appendix IA. 6 Shortcomings of Survey Methodology

In this online appendix, we discuss how we address two common shortcomings of survey methodology. One concern of survey design is social desirability bias (SDB), which refers to the tendency of participants to present themselves in a positive or socially desirable way (Haire 1950; Anderson 1978; Calder and Burnkrant 1977). To address this issue, we designed survey questions including both a direct and an indirect questioning technique. That is, rather than asking participants about their own behavior, we asked about their beliefs about general analyst behavior (referred to as general questions). Take our first question, for example. There, we asked if analysts agree with this statement: "In general, direct high-speed rails between brokerage houses and portfolio companies result in analysts conducting more site visits on their portfolio companies in person." Although this approach has been shown to mitigate SDB, one may be concerned that analysts incorrectly perceive the sensitivity of others to reductions in travel time. To alleviate this concern, we asked a second set of questions regarding analysts' perception of their own behavior (referred to as specific questions). For example, the specific question and scenario corresponding to the general question above is "The introduction of direct high-speed rail from Qingdao to Tangshan will increase the frequency with which *you* visit the portfolio company."

Another common issue pertaining to survey methodology is that participants might produce conflicting responses to questions that are worded slightly differently, referred to as description variance bias (Fischhoff 1991; Kuhberger 1998). To address this concern, we randomly broke down the remote participants into two subgroups (101 and 101 participants in each subgroup, respectively), with one subgroup of analysts receiving positively worded survey questions (original survey) and the other receiving negatively worded questions (alternative survey). An example of the alternative question corresponding to the example above is "The introduction of direct high-speed rail from Qingdao to Tangshan **will not** increase the frequency with which you visit the portfolio company." If the responses to these two sets of questions are consistent, we are confident our findings are unlikely to be driven by description variance bias. Fifty-five and 43 analysts responded to the original and alternative surveys, respectively. (See detailed original and alternative survey questions in Internet Appendix IA.7 and IA.8.) Their responses are highly consistent, as shown in Panels A and B of Internet Appendix IA.9 and depicted in Internet Appendix IA.10. To save space, we combine the responses from these two sets of surveys by inverting the responses to the alternative survey questions to align them with the original survey questions. For example, we invert the "strongly disagree" response in the alternative survey to "strongly agree" and "disagree" to "agree" and so on.

Internet Appendix IA. 7 Original Survey Questions

Please state whether you agree/disagree with the following statements:
"In general, direct high speed rails between brokerage houses and portfolio companies…

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
…. result in analysts conducting more site visits on their portfolio companies in person."	○	○	○	○	○
….result in analysts obtaining more firm-specific information on portfolio companies."	○	○	○	○	○
…result in analysts better understanding the strategies, operation, and performance."	○	○	○	○	○
…result in analysts better understanding the key challenges/issues their portfolio companies are facing."	○	○	○	○	○
…result in analysts better understanding the credibility of management, corporate culture and employees' morale."	○	○	○	○	○

For the next set of questions, please consider the hypothetical scenario below:

> Suppose you are an analyst based in Qingdao, Shandong and have followed a portfolio company in Tangshan, Hebei. Currently, the fastest way to travel between Qingdao and Tangshan is a regular rail taking ten to eleven hours (No airline route currently operates between the two cities). Suppose the Ministry of Railway (MOR) is planning to introduce direct high speed rail between Qingdao and Tangshan, which will substantially reduce travel time between the two locations.

Please state whether you agree/disagree with following statements:

1) The introduction of direct high speed rail from Qingdao to Tangshan will increase the frequency with which you visit the portfolio company.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

2) The introduction of direct high speed rail from Qingdao to Tangshan will increase your flexibility to visit the portfolio company when most useful.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

3) The introduction of direct high speed rail from Qingdao to Tangshan will help you better understand the credibility of management, current state of the company's strategies, operations, and performance, culture, and employment morale.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

4) The introduction of direct high speed rail from Qingdao to Tangshan will enable you to talk to non-management employees to gain more insights into the company's strategies, operations, culture, and employment morale.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

5) The introduction of direct high speed rail from Qingdao to Tangshan will help you make more accurate earnings forecasts for the portfolio company.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

6) The introduction of direct high speed rail from Qingdao to Tangshan will help you forecast the company’s long term growth.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

Analyst personal information:

1) The location of the brokerage house where you are employed:
   ________________Province
   ________________City

1. The number of firms you follow currently：
   A. < 3
   B. 3-8
   C. >8
2. Is there an HSR connecting you with any of your portfolio firm?
   A. Yes
   B. No
3. The average frequency of you visiting each of your portfolio firm:
   A. 0
   B. 1-3
   C. 3-5
   D. >5
4. Do you visit local portfolio firms more often that non-local firms on average?
   A. Yes
   B. No
5. Your portfolio firms concentrate in the following industry：
   A. Manufacturing industry
   B. High-tech industry
   C. Others

Internet Appendix IA. 8 Alternative Survey Questions

Below are the alternative survey questions worded negatively.

Please state whether you agree/disagree with the following statements:
"In general, direct high speed rails between brokerage houses and portfolio companies…

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
…. **will not** result in analysts conducting more site visits on their portfolio companies in person."	○	○	○	○	○
….**will not** result in analysts obtaining more firm-specific information on portfolio companies."	○	○	○	○	○
…**will not** result in analysts better understanding the strategies, operation, and performance."	○	○	○	○	○
…**will not** result in analysts better understanding the key challenges/issues their portfolio companies are facing."	○	○	○	○	○
…**will not** result in analysts better understanding the credibility of management, corporate culture and employees' morale."	○	○	○	○	○

For the next set of questions, please consider the hypothetical scenario below:

> Suppose you are an analyst based in Qingdao, Shandong and have followed a portfolio company in Tangshan, Hebei. Currently, the fastest way to travel between Qingdao and Tangshan is a regular rail taking ten to eleven hours (No airline route currently operates between the two cities). Suppose the Ministry of Railway (MOR) is planning to introduce direct high speed rail between Qingdao and Tangshan, which will substantially reduce travel time between the two locations.

Please state whether you agree/disagree with following statements:

7) The introduction of direct high speed rail from Qingdao to Tangshan **will not** increase the frequency with which you visit the portfolio company.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

8) The introduction of direct high speed rail from Qingdao to Tangshan **will not** increase your flexibility to visit the portfolio company when most useful.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

9) The introduction of direct high speed rail from Qingdao to Tangshan **will not** help you better understand the credibility of management, current state of the company's strategies, operations, and performance, culture, and employment morale.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

10) The introduction of direct high speed rail from Qingdao to Tangshan **will not** enable you to talk to non-management employees to gain more insights into the company's strategies, operations, culture, and employment morale.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

11) The introduction of direct high speed rail from Qingdao to Tangshan **will not** help you make more accurate earnings forecasts for the portfolio company.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

12) The introduction of direct high speed rail from Qingdao to Tangshan **will not** help you forecast the company’s long term growth.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
○	○	○	○	○

Analyst personal information:

2) The location of the brokerage house where you are employed:
________________Province
________________City

6. The number of firms you follow currently：
   D. < 3
   E. 3-8
   F. >8
7. Is there an HSR connecting you with any of your portfolio firm?
   C. Yes
   D. No
8. The average frequency of you visiting each of your portfolio firm:
   E. 0
   F. 1-3
   G. 3-5
   H. >5
9. Do you visit local portfolio firms more often that non-local firms on average?
   C. Yes
   D. No
10. Your portfolio firms concentrate in the following industry：
   D. Manufacturing industry
   E. High-tech industry
   F. Others

Internet Appendix IA.9 Survey Responses among Remote Respondents

Panels A and B present responses to the original and alternative survey questions, respectively, among remote respondents. If a respondent is employed by a broker located in a city other than Beijing, Shanghai, Shenzhen, and Guangzhou, she is deemed to be a remote respondent. For all questions, a 5-point scale is used (1 = *Strongly Disagree*, 5 = *Strongly Agree*). The % Agree column represents the percent of respondents who agreed or strongly agreed with the statement. A t-test is used to determine if the mean response is statistically different from the neutral midpoint of 3. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Remote Responses—Original Survey Questions (N=55)

Direct HSR connection results in analysts…	% Agree	Mean	St. Dev.
Conducting more site visits in person	98.18%	4.58***	0.53
Obtaining more firm-specific information	92.73%	4.42***	0.74
Better understanding strategies, operation, and performance	89.09%	4.38***	0.73
Better understanding key challenges/issues facing companies	83.64%	4.25***	0.91
Better understanding corporate culture and employees' morale	83.64%	4.29***	0.98
Increase frequency of visit	89.09%	4.33***	0.94
Increase flexibility to visit when most useful	96.36%	4.51***	0.77
Better understanding current strategies, operation, performance, corporate culture and employees' morale	83.64%	4.22***	0.96
Talk to non-management employees to gain more insights into the company's strategies, operations, culture, and employment morale	85.45%	4.29***	0.92
Make more accurate earnings forecasts	72.73%	4.00***	1.11
Help forecast the company's long-term growth	67.27%	3.78***	1.32

Panel B: Remote Responses—Alternative Survey Questions (N=43)

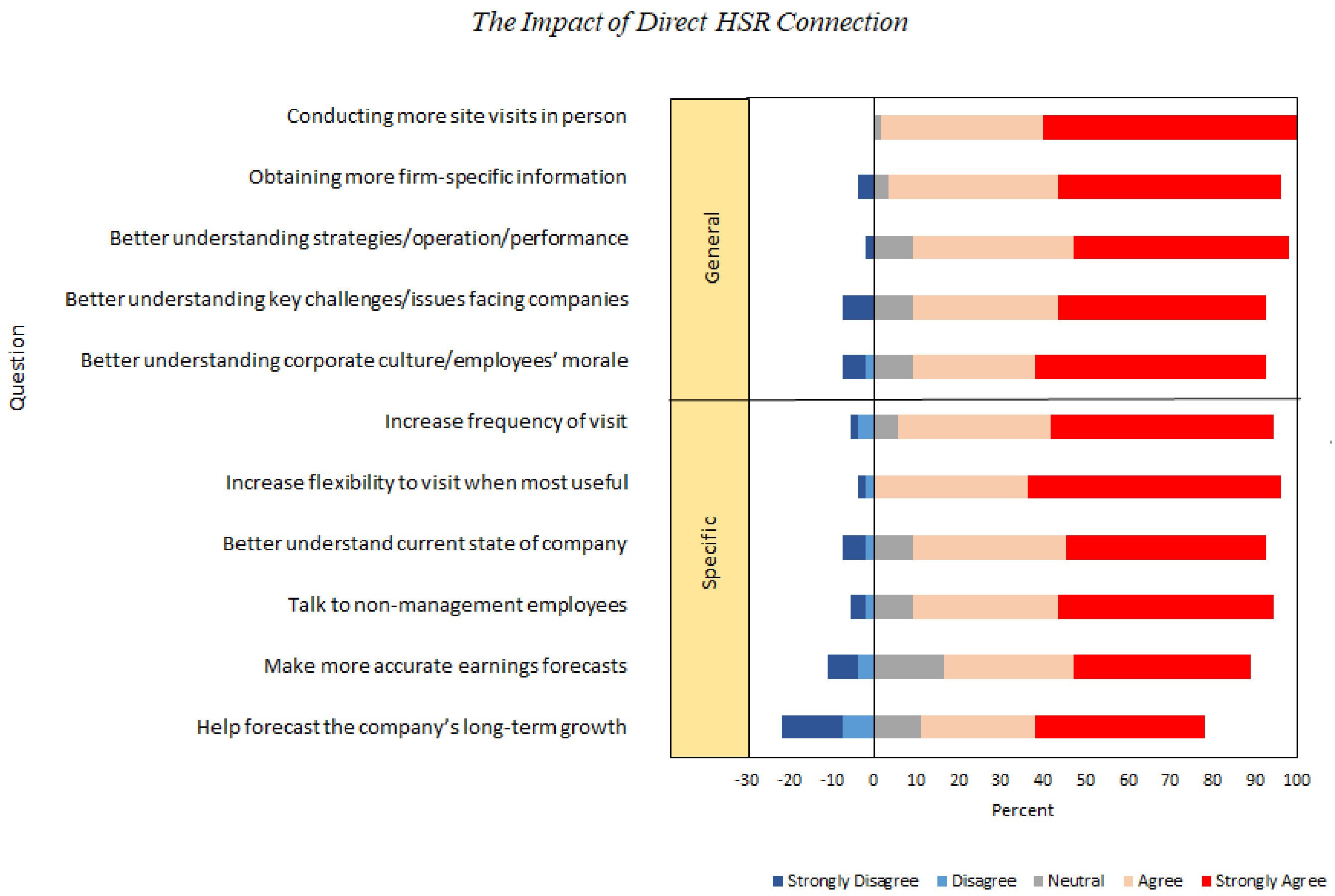
Direct HSR connection results in analysts…	% Agree	Mean	St. Dev.
NOT conducting more site visits in person	2.33%	1.40***	0.69
NOT obtaining more firm-specific information	2.33%	1.47***	0.7
NOT better understanding strategies, operation, and performance	4.65%	1.65***	0.9
NOT better understanding key challenges/issues facing companies	4.65%	1.56***	0.88
NOT better understanding corporate culture and employees' morale	2.33%	1.58***	0.79
NOT increase frequency of visit	2.33%	1.49***	0.77
NOT increase flexibility to visit when most useful	2.33%	1.47***	0.77
NOT better understanding current strategies, operation, performance, corporate culture, and employees' morale	0.00%	1.56***	0.67
NOT talk to non-management employees to gain more insights into the company's strategies, operations, culture, and employment morale	2.33%	1.53***	0.7
NOT make more accurate earnings forecasts	4.65%	1.58***	0.76
NOT help forecast the company's long-term growth	6.98%	1.65***	0.84

Internet Appendix IA.10 Response Distribution among Remote Analysts

Panel A: Survey among Remote Respondents Based on Original Survey Questions

This figure shows the distribution of responses to original survey questions among 55 remote respondents. If a respondent is employed by a broker located in a city other than Beijing, Shanghai, Shenzhen, and Guangzhou, she is deemed to be a remote respondent. The first five rows show the responses to the general questions, and the last six rows for the responses to the specific questions. On the horizontal axis, positive (negative) percentage refer to "strongly agree," "agree," and "neutral" ("disagree" and "strongly disagree") responses.

.



Internet Appendix IA.10 Response Distribution among Remote Analysts (Continued)

Panel B: Survey among Remote Respondents Based on Alternative Survey Questions

This figure shows the distribution of responses to alternative survey questions among 43 remote respondents. If a respondent is employed by a broker located in a city other than Beijing, Shanghai, Shenzhen, and Guangzhou, she is deemed to be a remote respondent. The first five rows show the responses to the general questions, and the last six rows for the responses to the specific questions. On the horizontal axis, positive (negative) percentages refer to "strongly agree," "agree," and "neutral" ("disagree" and "strongly disagree") responses.