Costly Information Acquisition and Investment Decisions: Quasi-Experimental Evidence

David Xiaoyu Xu†

This Version: August 2020

Abstract

Despite a vast theoretical literature that builds on costly information acquisition, there is no direct evidence on the role of information costs in investors' private information choices. Using a large sample of Chinese mutual fund managers' visits to firm headquarters and exploiting the introduction of high-speed railway as a quasi-natural experiment, I find substantial and quick responses to exogenous cost shocks: A one-standard-deviation travel time reduction increases site visits by 28% of the average frequency and semiannual stock trading profits by CNY 1.4 million at the fund family—firm pair level. These findings demonstrate investors' elasticity to information costs.

Keywords: Information Acquisition, Information Costs, Site Visits, Trading Profits.

JEL classifications: G11, G14, G23, D82, D83.

†University of Texas at Austin (Email: xyxu@mccombs.utexas.com). First version: August 2018. I am grateful to Aydogan Alti, Stephen Brown, Honghui Chen, Luis Goncalves-Pinto, Jennifer Huang, Erica Jiang, Travis Johnson, Xiang Kang, Jangwoo Lee, Tim Park, Clemens Sialm, and Sheridan Titman for helpful discussions and comments. I thank Lijin Fan for sharing data on the locations of fund families that have multiple offices. All remaining errors are my own.

I Introduction

Since Grossman and Stiglitz (1980), the paradigm of costly information acquisition has shaped our understanding of financial markets.[1] In this paradigm, investors choose to acquire private information based on a tradeoff between the costs of acquisition and the benefits of better-informed investment decisions. Despite the vast literature that builds on this tradeoff assumption, there is virtually no direct evidence from real-world investors. The goal of this study is to establish the empirical relevance of this microfoundation. To do so, I use direct observations on investors' information collection activities, and I estimate the extent to which investors respond to changes in information costs in a quasi-natural experiment.

The empirical setting of this study is based on Chinese mutual fund managers' visits to geographically dispersed corporate headquarters. The disclosure of private meetings between outside investors and corporate insiders provides a unique laboratory to analyze private information acquisition. Given the importance of proximity for site visits, I identify an exogenous shock to the cost of acquiring information: the introduction of high-speed rail that reduces travel times at the fund family–firm pair level. I then examine fund managers' responses by estimating the impact of this shock on their information choices and investment decisions.

The patterns of more than 100,000 site visits suggest that this setting reasonably fits theoretical models that feature costly information acquisition. First, fund managers' visits to a firm exhibit strong positive associations with their stock holdings, intraperiod trades, and trading profits, suggesting that private information and investment are jointly chosen.[2] Second, proximity appears to be an important dimension of information costs. Between fund family–firm pairs, travel times negatively correlate with the frequency of visits, portfolio weights,

[1]The investor's costly information acquisition has microfounded theories that shed light on various aspects of the market, such as equilibrium asset prices (Grossman and Stiglitz, 1980), managerial incentives (Holmström and Tirole, 1993), corporate financing choices (Subrahmanyam and Titman, 1999), and asset management industry (Gârleanu and Pedersen, 2018).

[2]Although visits can generate both good and bad posterior beliefs about a firm, mutual funds' short-sale constraints largely prohibit fund managers from trading on negative signals.

intraperiod trades, and trading profits. These negative correlations hold even after controlling for geographical distances, which provides suggestive evidence for the relationship between information costs and investor decisions.

The variation in travel times between fund family–firm pairs, however, does not identify the causal effects of travel times on site visits and stock trades. The empirical challenge is that travel time, as one of the determinants of fund managers' decisions, might correlate with other determinants that are unobservable to the econometrician. For example, fund managers not only face lower costs of acquiring information about nearby firms, but may also have better prior knowledge about these firms. Such knowledge is likely to affect both learning and trading decisions, giving rise to spurious correlations between travel times and these outcomes.

To address this endogeneity problem and isolate the causal effects, I exploit exogenous within pair variation in travel times induced by the introduction of high-speed rail. The growth of the rail network reduces travel times for several subsets of fund family–firm pairs in different periods. I define the introduction of a new rail line as a treatment on such pairs, which are assigned to the *treated* group. In contrast, the rail network does not affect fund family–firm pairs for which driving or air travel is faster. These pairs are assigned to the *control* group.

Three facts help ensure that the treatment is orthogonal to pair-specific variables other than travel times (e.g., fund managers' prior beliefs). First, the rail network was solely designed by the Chinese government, whose decisions are independent of mutual fund investing activities at the pair level. Second, each rail line requires several years of construction before the predetermined introduction event occurs, so the treatment timing is unlikely to coincide with omitted time-varying pair-specific shocks. Third, the high-speed rail serves passengers but does not affect freight transport, which rules out the possibility that the treatment correlates with the outcomes through supply chains or product markets.

Importantly, my empirical setting allows me to address a key identification challenge in this quasi-experimental design; namely, that the introduction of rail lines might correlate with firm fundamentals.[3] Given that all mutual funds have similar access to visiting and trading every public firm, I can use within firm-by-time variation in the treatment status. My identification strategy compares the frequency of visits to (and investment decisions on) the same firm by fund families for which travel times are reduced and by other fund families, both before and after the treatment. Since firm-level shocks should affect the treated and control pairs similarly, I identify the effects of travel time reductions using their differential responses to the treatment.

I implement this identification strategy in a difference-in-differences framework using high-dimensional fixed effects. On average, the introduction of a high-speed rail line reduces pre-treatment travel times for treated pairs by 90% of a standard deviation. My estimates show that this reduction increases the frequency of site visits by 25% relative to the unconditional average (5% of a standard deviation). Meanwhile, the reduction also increases trading profits by CNY 1.3 million (4% of a standard deviation) during a 6-month period.[4] These results indicate that fund managers respond substantially to reductions in information costs by acquiring more private information.

I then examine the time-series and cross-sectional implications of travel time reductions. A dynamic test shows that the treatment effects are not driven by pre-existing trends: These effects emerge immediately after new rail lines start service, and the effects persist over multiple years. Across the treated pairs, the treatment effects are stronger for pairs with larger travel time reductions. These effects largely come from relatively distant pairs for which traveling between addresses used to be costly, and firms in manufacturing industries perhaps because more soft information about tangible assets is effectively collected on site. In a placebo test, I

[3]For example, better transport service could improve investment opportunities for local firms. Also, the locations of rail stations could be selected based on local economic prospects. In these cases, the finding of positive effects would be spuriously driven by omitted firm-level shocks if fund managers visit growing firms more frequently.

[4]The causal effect of travel times on trading profits should not be interpreted as merely driven by fund managers' visits because travel times also affect other pair-level activities that are unobservable to the econometrician. Such activities include corporate insiders' visits to financial hub cities (e.g., non-deal roadshows) and fund managers' other forms of information acquisition on the firm.

find no effect on fund managers' participation in remote meetings or conference calls. Results of these tests provide additional support for the interpretation that the estimated effects are driven by changes in travel times.

This study is closely related to two literatures. The first is a finance literature that infers the existence of private information from mutual fund portfolios. Early papers in this literature show a negative relationship between geographical distance and a stock's portfolio weight and future return (Coval and Moskowitz, 1999, 2001).[5] Subsequent papers find that fund managers have access to information from corporate board members with shared education networks (Cohen, Frazzini, and Malloy, 2008), banks within the same financial group (Massa and Rehman, 2008), and other fund managers in the same city (Hong, Kubik, and Stein, 2005) and neighborhood (Pool, Stoffman, and Yonker, 2015). The second is a growing literature that studies the implications of transportation infrastructure changes. This literature generally finds that the introduction of new airline routes has an impact on economic outcomes, including investment and productivity of corporate plants (Giroud, 2013), venture capitalists' investment success (Bernstein, Giroud, and Townsend, 2016), and institutional investors' stock holdings (Ellis, Madureira, and Underwood 2019, Da et al. 2020).

While existing papers mostly focus on ex post investment decisions, this study differs in that it examines mutual fund managers' ex ante information choices and their responses to information cost shocks in a setting that is closer to related theoretical models.[6] Beyond documenting the impact of travel times on investment, it directly quantifies the cost elasticity of investor demand for private information and provides the first direct evidence for a large number of theories on costly information acquisition. In addition, this study's identification

[5] In a contemporaneous paper, Chen et al. (2019) document that Chinese mutual funds exhibit a similar local preference in both portfolio holdings and site visits.

[6] Travel times affect investor trips as well as trips by other individuals (e.g., corporate insiders), so portfolio changes do not necessarily reflect investors' visits. Even if investor trips are observable, in general, institutional investors visit corporations for both informed trading (Ellis, Madureira, and Underwood, 2019) and monitoring (Da et al., 2020). In China, mutual funds not only have observable visits, but are also minority shareholders of public firms and typically passive in corporate governance, thus providing an ideal setting for studying trading-motivated information acquisition. See the Appendix for related institutional details.

strategy requires a weaker assumption on the exclusion restriction and complements existing empirical designs based on commercial airlines.

This study also contributes to empirical research on investor information acquisition. The majority of this literature uses Google search volume (e.g., Da, Engelberg, and Gao 2011 and Drake, Roulstone, and Thornock 2012) and web traffic to the SEC's EDGAR filings (e.g., Chen et al. 2017; Gallagher et al. 2018; Chen, Kelly, and Wu 2018; Crane, Crotty, and Umar 2018) as proxies for the processing of public information. The literature also identifies two ways through which hedge funds acquire private information: hiring lobbyists (Gao and Huang, 2016) and sending FDA–FOIA requests (Gargano, Rossi, and Wermers, 2017). Moreover, research based on data from the same disclosure rule in China finds that site meetings are related to stock prices (Cheng, Wang, and Wang, 2017), insider trades (Bowen et al., 2018), and sell-side analyst forecast accuracy (Cheng, Du, Wang, and Wang 2016, Han, Kong, and Liu 2018, Dong et al. 2019). These papers do not study the implications of investors' information costs.

The remainder of this paper proceeds as follows. Section II explains the data sources, the sample and the empirical measures. Section III investigates the patterns of site visits and evaluates travel time as a measure of information costs. Section IV introduces the quasi-natural experiment and presents estimation results. Section V concludes.

II Data

The data used in this study come from three main sources. From the China Stock Market & Accounting Research (CSMAR) database, I obtain historical firm information, stock returns, mutual fund portfolio holdings, and cumulative intraperiod stock trades. The historical addresses of firm headquarters and mutual fund families are manually verified based on raw data from this database. Travel times between mutual fund offices and firm headquarters are computed based on Web APIs of two travel navigation service providers. Mutual fund site visit records are hand collected from mandatory disclosure reports of firms' investor relations activities. The Appendix provides greater details of travel time computation and the hand collection of private meetings data.

A Sample Construction

A.1 Firms

I begin with all 3,500 firms that are publicly traded on China's two major stock exchanges: the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE). A firm is included if its stock is ever listed during 2008–2017 on one of the following trading boards: the main boards of the two exchanges, the SZSE's Growth Enterprise Market (GEM), or the Small/Medium Enterprise (SME) boards. These stocks account for more than 95% of mutual fund equity security holdings. Based on the China Securities Regulatory Commission (CSRC) industry classification, 82 firms belong in the financial category. I exclude these firms because site visits are less relevant for firm-specific information.[7] Next, I trace back the annual history of each firm, and determine whether a firm ever experienced a material office move during the sample period. There are 295 movers, and excluding them further removes 287 firms from the sample.[8] The resulting sample consists of 3,131 unique firms.

A.2 Mutual Funds and Fund Families

I begin with a survivor bias free set of 5,660 unique mutual funds, then I exclude funds of funds (CategoryID=S0605), exchange-traded funds (IsETF=1), and index funds (IsIndex=1). I also remove funds that are categorized as passive investment vehicles (IsActiveOrPassive=2).

[7] Exclusion of financial and mover firms are done after computing portfolio weights.

[8] Since there are erroneous records in firm headquarter office zip codes, this step is achieved with careful visual inspection and manual correction. Out of these 295 firms, 148 firms experience office moves because they are acquired by private firms (reverse mergers) located in different cities. In almost all of these cases, the acquirer firms inherit the target firms' stock ticker symbols.

This procedure results in a list of 4,882 mutual funds, and these funds' portfolio holdings and trades data are used in this paper.

I include all 114 China-domiciled mutual fund families that report portfolio stock holdings during 2008–2017. A fund family's office location is defined as the address of the building where portfolio managers and buy-side analysts work. Although several fund families are registered in other cities for tax reasons, all mutual fund families' offices are located in the central business districts (CBDs) of one of the four metropolitan cities in China: Beijing, Shanghai, Guangzhou, and Shenzhen.[9]

A.3 Portfolio Holdings and Cumulative Trades

Since 2004, China-domiciled mutual funds are required to fully disclose their portfolio equity holdings every six months. I begin with all 1,750,974 domestic stock holding records from semiannual and annual reports between 2008–2017, and remove any holdings that are labeled by the fund as index investment (InvestmentType=2). Next, I use the sample stock and sample mutual fund lists to screen for holdings of non-passive funds. After this filter, 972,353 fund–stock records for 2,790 unique mutual funds remain.[10] To eliminate private placement stock holdings, I further drop 29,285 records in which the holding date is before the stock's IPO date.

The disclosure filings also provide 6-month stock buy and sell values (separately, in CNY). Mutual funds are required to report cumulative cash flows generated from all material stock trades: Whenever a fund's intraperiod purchase volume of a stock exceeds 2% of period-beginning fund TNA, the fund discloses the cumulative amount of money spent in buying this stock. If a fund has fewer than 20 stocks that satisfy this criteria during a period, then the fund discloses the cumulative amount of money spent in buying each of the top-20 stocks in

[9]Several mutual fund families' office locations differ from their headquarters locations. For these fund families, I use the locations of their the actual offices of portfolio managers and equity analysts.

[10]I include all active stock holdings of equity, balanced, and bond funds, regardless of whether they are open-end or closed-end.

terms of purchase volume. The disclosure requirement for the cumulative amount of money received from intraperiod stock sales is the same.[11] These cumulative amounts provide rich information on short-term trading activities. Similar to portfolio holdings, I obtain 1,623,559 cumulative intraperiod trading records during the sample period to construct the trading profit measure.

A.4 Panel Structure

I take a Cartesian product of the three sets of unique identifiers (i.e., firms, fund families, and semi-year dates during 2008–2017) to set up a panel dataset. Hence, each observation in this dataset is identified by a pair and a semi-year, where each pair consists of a fund family and a firm. I remove an observation if the firm is not yet listed on an exchange, or has been de-listed, or if the fund family does not appear in holdings data (i.e., it is not established) at the end of the period. This results in a sample with 3,347,853 observations.

Table I summarizes the composition of this sample. The dataset is an unbalanced panel due to the quick growth in the numbers of both publicly traded firms and mutual fund families. As shown in Panel A, more firms became listed on SZSE than on SSE, and the total number of firms more than doubled during the 10-year sample period. The number of mutual fund families located in Beijing and Shenzhen increased the most, and the total number of fund families nearly doubled. The contemporaneous increase in both institutional investors and firms led to even faster growth in the number of pairs.

Panel B shows the industry category distribution of the sample firms. Consistent with China's economic growth path after joining the World Trade Organization (WTO) in 2001, more than half (66.0%) of these firms belong to manufacturing industries. The second and third largest industry categories are information technology and wholesale & retail, followed by construction & utilities. Site visit is a desirable tool for acquiring information about these

[11] On average, a fund reports 6-month cumulative buy and sell values for 36.5 and 36.2 stocks, respectively.

firms because of the tangibility of their assets.

B Variables

Given the data structure of this sample, the majority of variables are defined at the fund family–firm pair level. The underlying assumption is that site visits conducted by employees of a fund family generate information about a firm for multiple funds within the family. This assumption is justified by the fact that it is common practice that a portfolio manager manages multiple funds and a buy-side analyst supports multiple portfolio managers of the fund family. All CNY-valued variables are expressed in 2006 CNY after adjusting for inflation.

B.1 Geographical Distance

Following Coval and Moskowitz (1999), I compute the distance between each fund family i's address and each firm f's headquarters office address based on their latitudes and longitudes:

$$Distance_{i,f} = 2\pi r \times \arccos\left(\Pi_{i,f} + \Theta_{i,f} + \Phi_{i,f}\right)/360, \tag{1}$$

where r is the radius of the earth, and other variables are $\Pi_{i,f} = \cos(lat_i)\cos(lon_i)\cos(lat_f)\cos(lon_f)$, $\Theta_{i,f} = \cos(lat_i)\sin(lon_i)\cos(lat_f)\sin(lon_f)$, and $\Phi_{i,f} = \sin(lat_i)\sin(lat_f)$. This distance measure reflects the length of a "frictionless" trip between two points on the surface of the earth, but it does not account for geographical features or means of transport.

B.2 Travel Time

Using Web APIs provided by commercial navigation applications, I develop an algorithm to compute the travel time between the addresses of each fund family and each firm headquarters. I define travel time as the estimated number of minutes of travel, based on optimized combinations of transport segments (e.g., driving, trains, and flights). Specifically, I generate three itineraries for each pairing of a trip's origin and destination, and each itinerary represents one feasible travel plan. The first is a *car-based* travel plan, and *DrivingTime* is the time duration for a one-way trip using only a car. The second is a *train-based* travel plan, and the third is a *flight-based* travel plan.[12] For the second and third plans, I force the navigation planner to prioritize the corresponding means of transport whenever they are available. Given the three travel time estimates for each origin–destination pair, I assign the shortest time among them as the value of *TravelTime*.

B.3 Number of Site Visits

Since 2004, the Shenzhen Stock Exchange mandates the disclosure of private meetings between firm management and outside investors. I hand-collect mutual fund corporate site visit records from mandatory disclosure filings for all SZSE-listed firms between 2008 and 2017. For each private meeting, a typical report discloses the date and location of the meeting, as well as the names of the attendees and their respective employers.[13] In addition, the report classifies the meeting into various types, including site visits and conference calls.

I begin with a dataset of 133,785 visitor–firm–event records that involve mutual fund employees, and I aggregate these records to 95,483 fund family–firm–date observations. Next, I divide all meeting events into two groups based on whether they are held at the firm's headquarters offices or elsewhere. Then, I aggregate each fund family's site visits and remote meetings at each firm during each semi-year period to obtain a measure that is consistent with fund investment disclosure frequency. The final site visit dataset contains 81,143 pair–semi-year observations. The remote meetings dataset includes 6,231 observations. I match these observations to the panel dataset, and I assign zero values for the remaining observations that experience no meeting event.

[12]The computed travel times partly depend on the distance between a firm's headquarters and the nearest airport (or rail station). My travel time computation algorithm accounts for these details.

[13]A limitation is that in most cases the employer is a fund family, so the data do not allow me to link meeting participants to individual mutual funds.

B.4 Active Portfolio Weight

At the end of each period, I aggregate the portfolio holdings of each stock over all sibling mutual funds of each fund family to construct an active portfolio weight measure. This market-adjusted portfolio weight is defined as

$$ActiveWeight_{i,f,t} = \left| \frac{Holding_{i,f,t}}{\sum_f Holding_{i,f,t}} - \frac{MktCap_{f,t}}{\sum_f MktCap_{f,t}} \right|, \tag{2}$$

where $Holding_{i,f,t}$ is the sum of the market value of firm f's stock reported to be held in portfolios by all sibling funds of fund family i at the end of period t, and $MktCap_{f,t}$ is firm f's market capitalization measured based on the total number of tradable shares at the end of period t.

B.5 Trading Profit

Similar to Irvine, Lipson, and Puckett (2006) and Puckett and Yan (2011), I combine mutual fund end-of-period stock holdings and intraperiod cumulative stock trades data to construct a pair-level trading profits measure[14]:

$$Profit_{i,f,t-1\to t} = Holding_{i,f,t} + Sell_{i,f,t-1\to t} - Buy_{i,f,t-1\to t} - Holding_{i,f,t-1},$$

where $Buy_{i,f,t-1\to t}$ (or $Sell_{i,f,t-1\to t}$) is the cumulative absolute CNY values that fund family i pays for purchasing (or receives from selling) firm f's stock between the ends of period $t-1$ and period t. To adjust for cash dividends, I add back dividend payments based on the average number of shares held by a fund family at the beginning and end of a period.[15]

This measure differs from conventional return-based performance metrics because it re-

[14] I refer to this measure as "trading profit" and "investment profit" interchangeably. Note that this measure differs from the fund family's own profit.

[15] The exact number of shares held on the date when the firm pays out dividends is not observable. However, ignoring dividends, or adjusting for dividends in different ways, has no material influence on my results.

flects both market price changes of unaltered stock holdings and cash flows from intraperiod trades. Although the timing and magnitude of each stock trade are still unobservable, this measure effectively captures the performance of fund trading decisions. As will become evident, capturing the contribution of these unobservable trades to performance is critical for testing private information because Chinese equity funds have high portfolio turnover and portfolio snapshots are only available every six months.[16] The measure is not adjusted for risks or characteristics, but this limitation does not bias the results because trading profits from the same stock during the same period will be compared between different investors.

C Summary Statistics

Table II summarizes the distributions of the main time-varying variables. Panel A shows firm characteristics by aggregating observations to the firm–semi-year level. The average firm has a market capitalization of CNY 9.0 billion, with 5.9% of tradable shares outstanding held by 9.4 mutual fund families. The average SZSE-listed firm receives three site visits by mutual fund managers during a 6-month period, while more than half of firm–semi-year observations do not have a mutual fund visit.[17] In terms of stock returns, these firms show a wide range variation.

Panel B presents the sample from the perspective of mutual fund families. Meeting with firm management seems to be an important activity for mutual funds: On average, during each period, a fund family's employees make 49.5 trips to visit SZSE-listed firms and participate in 3.9 remote meetings. Compared with the universe of publicly-traded firms, Chinese mutual funds' stock portfolios are highly concentrated: The average fund family holds only 261.1 stocks. The sizes of these portfolios also vary substantially, from CNY 1.8 billion at the 25th percentile to CNY 20.1 billion at the 75th percentile.

[16]See Table A.1 in the Appendix for a summary of unobservable mutual fund intraperiod trading activities.

[17]Mutual funds account for approximately 30% of all visitors during 2012-2017, and more than 70% of SZSE-listed firms experience site visits from different types of visitors, including mutual funds, hedge funds and sell-side analysts.

Finally, Panel C shows variables at the pair–semi-year level. Site visits occur in fewer than 5% of pairs with SZSE-listed firms in a period, and a firm's stock is held by the fund family only in 11.6% of pairs. This ratio is almost the same for intraperiod stock trades. On average, fund families' investment profit is modest. This is related to the fact that market portfolio return is approximately -25% during 2008–2017. However, there is a wide dispersion of pair-level trading profits on both sides.

III Site Visits and Stock Investment

This section explores mutual fund managers' information acquisition and investment decisions. I first combine site visits and mutual fund stock investment to examine the collection and utilization of private information. I then evaluate travel time as a measure of information costs.

A Background

Since its beginning in 1998, the Chinese mutual fund industry has grown quickly along with the Chinese economy. According to the Asset Management Association of China, by the end of 2017, size of the total mutual fund assets under management reached CNY 11.7 trillion, out of which CNY 2.7 trillion are managed by equity and hybrid funds. Similar to the US market, common stocks of publicly traded domestic firms constitute one of the major asset classes held by Chinese mutual funds.

Mutual fund managers and analysts are frequent travellers.[18] Even in today's digital era, site investigations and face-to-face communication are still useful for collecting firm-specific information that is not publicly available.

[18] In *Beating the Street*, Peter Lynch writes "*My visits with companies, either at our place or at their places or at investment seminars, also had escalated from 214 in 1980 to 330 in 1982, 489 in 1983, back down to 411 in 1984, 463 in 1985, and 570 in 1986. If this kept up, I figured I'd be seeing an average of two companies a day in person, including Sundays and holidays.*"

Firms' geographic locations are important for site visits because visitors must be physically present. Figure I plots the headquarters locations for all sample firms and financial hub cities where mutual fund families are located. Although more firms are located in better-developed regions (e.g., the Yangtze River delta in the east, and the Pearl River delta in the south), overall, these firms are dispersed across all provinces of China. Large geographical dispersion generates variation in travel-related costs for fund managers. The degrees of dispersion are similar between the two groups of firms listed on the two stock exchanges. Moreover, since distances are long among the four financial hub cities, then travel time between a given firm and different fund families can vary substantially. These facts provide the variation for discovering the effects of travel times if they are present.

B Site Visits and Investment Decisions

Mutual fund managers are well compensated, and their business trips are costly for fund families. Can the large number of site visits observed in data be justified by the information acquired from these trips? In Table III, I report the results of regressions that explore the relationship between site visits and fund investment decisions.[19] In these regressions, I control for pair fixed effects so that the estimates come from within pair variation. I also include firm-by-time fixed effects and fund family-by-time fixed effects to absorb any time-varying effects at the firm and fund family levels, thus ensuring that pair-level quantities are fairly compared across pairs.

Column (1) shows that holding the firm's stock at the end of the previous period is associated with 0.017 more visits during the current period. Column (2) replaces the holding dummy with portfolio active weight and shows that 1 percentage more holding at the end of the previous period is associated with a similar number of visits during the current period. Re-

[19] In all regression analyses, I do not restrict the sample to observations for which the outcome variable has nonzero values. Hence, the estimates depend on whether the fund family acts on a firm (extensive margin) and the action's magnitude (intensive margin) in a period.

sults in Columns (3) and (4) indicate that 1 visit during the current period is associated with 5 basis points of larger active weight and 8.4% of higher likelihood of holding the firm's stock at the end of the current period, respectively. These positive correlations arise from the fact that Chinese mutual funds are short-sale constrained and for this reason, positive signals are more likely to affect investment than negative signals. Column (5) shows that 1 visit corresponds to 5.9% higher likelihood of trading the firm's stock during the current period.[20] All point estimates in Table III are highly statistically significant, suggesting that fund managers jointly choose site visits and investment decisions.

In Table IV, I examine the importance of site visits and intraperiod trades on mutual fund investment performance. Panel A reports average pair-level trading profits by whether a site visit and an intraperiod trade occurs. When there is neither visit nor trade, average investment profit is close to zero. Mutual funds realize more profits when they visit firms or trade stocks during the period. The average profit is especially high when both site visit and intraperiod trade occur, which suggests that the short-lived private information acquired during site visits improves investment performance.

Panel B reports results from regressing trading profits on the number of site visits and whether intraperiod trade occurs. Trading profits depend on how much a fund family invests in a stock and whether the investment idea is good, both of which are endogenously driven by the fund managers' information. To make the comparison of CNY-valued trading profits meaningful, in these regressions I include the same fixed effects as those in Table III.[21] Results in Columns (1) to (3) indicate that on average, a fund family realizes CNY 1.2 million additional profits from a stock when fund managers visit the firm and CNY 2.1 million additional profits if intraperiod trades occur. Consistent with Panel A, these results indicate that mutual

[20]Although the intraperiod trades data allow us to measure whether a stock purchase or sale occurs, the number of shares traded are still unobservable.

[21]For example, the firm-by-time fixed effects ensure that the estimated differences reflect variation in investment profits from the same stock during the same period, thus having exactly the same market capitalization and risk conditional on public information.

funds acquire private information from site visits and realize investment profits mainly from short-term stock trades.

C Travel Time as Information Cost

This subsection evaluates travel time as a measure of the cost of acquiring information. A well known empirical fact is that mutual funds overweight firms headquartered nearby (Coval and Moskowitz, 1999). If such geographical patterns are driven by travel-related information costs, then travel time should explain these patterns better than geographical distance. To examine this conjecture, I compare the two proximity measures in terms of their empirical relation to fund managers' decisions.

Figure II plots travel time and driving time as functions of geographical distance for all origin–destination pairs in the sample. These two time measures largely overlap when the distance is less than 300 kilometers, where travel by car tends to be the most efficient means of transport. Clearly, as the distance increases, travel time flattens while driving time increases linearly. The concavity of travel time as a function of distance is due to the efficiency of trains and airplanes when traveling over longer distances.

Beyond distance, travel time is affected by mountains, rivers, and other landforms. The location of a firm's headquarters relative to the nearest airports and railway stations also has a considerable impact on travel time. In Figure II, the vertical variation reflects the effect of these factors on travel time given the same distance. For some distant pairs, travel time is close to driving time due to the lack of other means of passenger transport. Such variation allows for disentangling the effect of travel time from the effect of distance.

In Table V, I estimate the relations between the two proximity measures (i.e., travel time and distance) and fund managers' site visits, portfolio weights, intraperiod trades, and trading profits. In these regressions, the inclusion of firm-by-time fixed effects and fund family-by-time fixed effects ensure that the coefficients are estimated using only between pair variation in proximity and the outcomes.

When either travel time or geographical distance is the only regressor, the estimated coefficients are negative and statistically significant. In Panel A, the number of site visits and active portfolio weights decrease by 0.007 times and 0.1 basis points for a 1-hour increase in travel time. The magnitudes of their decreases for a 1,000-kilometer increase in distance are 0.022 times and 0.33 basis points, respectively. When both travel time and distance are included as regressors, as in Columns (3) and (6), the coefficients on travel time remain significant with moderately smaller magnitudes, while the coefficients on distance become small and insignificant.

In Panel B, the correlations between the proximity measures and stock trades (and trading profits) are also negative and statistically significant. In terms of economic magnitude, a 1-hour increase in travel time is associated with 0.1% lower likelihood of stock trade and CNY 0.07 million less trading profits. Conditional on distance, the coefficients on travel time remain similar, but conditional on travel time the coefficients on distance become small and insignificant. Overall, the results in Table V support the conjecture that travel time is a better measure of travel-related information costs than geographical distance. These negative correlations also provide suggestive evidence for the effects of information costs on mutual fund decisions.

IV A Quasi–Natural Experiment

This section proceeds to estimate the causal effects of information costs on information acquisition activities and investment decisions. The identification challenge is to isolate the effects of information costs from the effects of pair-level variables that also correlate with proximity. For example, the OLS estimates are biased if between pair variation in travel times correlates with fund managers' prior knowledge about nearby firms.

To address this endogeneity problem, an ideal experiment would force the travel times to change for a random subset of fund family–firm pairs, leaving all other pairs unchanged. Although such an experiment is difficult to implement in the real world, I approximate it by exploiting the Chinese government's long-term investment in one particular public transport technology.

A China Railway High-Speed

Since 2008, China has experienced a phenomenal expansion of its high-speed railway network, which is named China Railway high-speed (CRH). From nearly zero in 2008, the CRH network quickly grew to comprise more than 60% of the global total length of high-speed rail operating at the end of 2017. Relative to other means of passenger transport (i.e., cars and airplanes), CRH exhibited a remarkable annual growth rate during the sample period.[22]

The introduction of new CRH lines provides useful within pair variation in travel time, primarily because of its door-to-door speed advantage. In regular operation, CRH trains travel at speeds between 250 km/hour and 350 km/hour. When the distance is medium, air travel's speed advantage is offset by the time required to drive to and from the airports, which are typically located in suburban areas. The slow security process in the departure airport also adds to the travel time. If a firm is located close to a CRH station but far from the nearest airport, the introduction of a new CRH line can lead to a dramatic reduction in travel time.[23]

For a subset of pairs in the sample, the introduction of CRH lines reduces travel times and makes it less costly for fund managers to visit the firms. To exploit such within pair variation in travel time induced by new CRH lines, I define a pair as *treated* if the train-based travel plan is faster than the second fastest plan by at least 30 minutes, and if at least one segment of this

[22] See Figure A.2 and Figure A.3 in the Appendix for the development of this public transport technology.

[23] The timeliness of CRH adds additional economic significance to the treatment. Air travellers in China frequently suffer from delays and cancellations, which particularly affect passengers who have tight meeting schedules.

optimal travel plan involves CRH trains.[24] Under these criteria, 13,190 pairs are ever treated. Figure III shows the distribution of geographical distances for all pairs and the distribution of travel times for pairs that are ever treated. The distances of the majority of treated pairs are between 300 and 1,000 km. This fact is consistent with the high-speed rail's speed advantage over medium distances relative to driving or flying. For treated pairs, the average travel time reduction is 61.5 minutes (the median is 55.7 minutes) for a one-way trip, which is around 90% of a standard deviation of pre-treatment travel time (Table VI).

To determine the timing of the treatment events, I manually collect the date when each segment of every CRH line started its passenger service from historical public news.[25] Since all events occurred between June and December of the corresponding years, I group all treated pairs into eight treatment cohorts from 2009–2016. In the regressions, I do not require that treated pairs existed before or after the treatment events. However, results are similar if I use a treatment window (i.e., a particular number of periods before and after events) to select observations for treated pairs.

B Econometric Specification

The baseline specification in this empirical design is a generalized difference-in-differences (DiD) approach with multiple treatment events. I estimate equation

$$Outcome_{i,f,t} = \beta \times Treatment_{i,f,t} + \alpha_i \times \alpha_f + \alpha_f \times \delta_t + \alpha_i \times \delta_t + \varepsilon_{i,f,t}, \tag{3}$$

where $Treatment_{i,f,t}$ is a dummy variable that equals 1 if the high-speed rail line that reduces travel time between fund family i and firm f is in service during period t. The main outcome

[24] To mitigate the noise in travel time computation, I exclude pairs for which the travel times of the train-based plan do not differ from previous travel times for more than 30 minutes. An itinerary segment is determined as involving CRH trains if the train's ID number begins with letter *G*, *D* or *C*.

[25] If more than one CRH segments are involved in a travel plan, I use the last connected segment to determine the treatment timing, because this is the time when the travel plan became feasible.

variables are the number of site visits and the amount of trading profits.[26] The difference-in-differences estimator $\hat{\beta}$ captures the treatment effects of travel time reductions on the outcome variables.

The estimation is facilitated by high-dimensional fixed effects. In Equation (3), α_i and α_f denote unique identifiers for fund families and firms, and δ_t denotes unique semi-year dates. The first interaction term, $\alpha_i \times \alpha_f$, denotes pair fixed effects that absorb the impact of pair-specific time-invariant heterogeneities such as geographical distance. The second interaction term, $\alpha_f \times \delta_t$, denotes firm-by-time fixed effects. Such fixed effects control for any firm-level economic shocks and ensure that the outcomes are compared between fund families given the same firm and period. I use a third group of fixed effects, $\alpha_i \times \delta_t$, to control for fund family-level shocks that affect the outcomes, such as the fund family's asset size and manager skills.

This specification presents a high empirical hurdle. Important determinants of the outcome variables, such as firm fundamentals, are differenced out. Essentially, this specification compares the changes in the number of site visits to (trading profits from the stock of) the same firm before and after the treatment, between fund families whose travel times are reduced and fund families whose travel times remain constant. Hence, the inclusion of the firm-by-time fixed effects is crucial and ensures that appropriate within variation is used in the estimation.[27]

C Identification

My identification strategy exploits differential responses to the staggered introductions of high-speed rail lines that reduce the travel times between firms and fund families. The broad

[26] I use the CNY-valued measure for trading profits to capture the benefits of information. Others equal, two fund families that have symmetric information about a firm should generate the same expected amount of profits from the firm's stock during a given period. In the Appendix, I report results using scaled versions of this measure.

[27] Despite that the treated and control groups are uniquely defined at the pair level, firm-level confounders can still lead to spurious results because the control pairs outnumber the treated pairs.

geographical dispersion of fund families and firms ensures that, when a CRH line begins service, it provides a new optimal travel plan for only a subset of pairs. As a result, the same firm can appear in both the treated and control pairs, and different treated pairs can receive the treatment in different periods.[28] Without these desirable features and the rich variation they generate, I could not identify the interested effects even if they existed.

The identifying assumption is that, conditional on travel times and the fixed effects, the CRH introduction events are uncorrelated with the outcomes. This assumption is weak and plausible in my empirical setting for the following reasons. First, all CRH lines are entirely designed and financed by the central government of China, whose decisions are unlikely to be related to pair-specific investing activities. Hence, the assignment of treatment should be conditionally uncorrelated with potential outcomes. Second, the construction of the CRH lines typically requires three to five years before predetermined introduction events.[29] This fact largely alleviates the concern that the timing of treatment might correlate with pair-level omitted variables such as fund managers' prior beliefs. Moreover, the CRH network serves passengers but does not affect freight transport, so the treatment should not correlate with the outcomes through supply chains or product markets. Under this identifying assumption, the difference in fund managers' responses to the treatment captures the causal effects of travel time reductions.

A limitation of this empirical design is that the average effect of each site visit cannot be identified using exogenous variation in travel times. This is because travel times also affect other pair-level activities that are unobservable to the econometrician, such as corporate insiders' visits to financial hub cities (e.g., non-deal roadshows) and fund managers' other forms of information acquisition on the firm. For the same reason, the treatment effect on trading profits should not be interpreted as merely driven by fund managers' visits. That said, the

[28]See Figure IV for a visualization of treated pairs, divided into four panels according to the financial hub where paired mutual fund families are located. This figure illustrates the features discussed above.

[29]Future events are publicly announced before the actual travel time reductions take place. See Bullock, Sondhi, and Amos (2009, p. 75) for a summary of the predetermined CRH construction plans.

results are still informative about the effects of travel times on fund managers' information choices.

D Results

As discussed earlier, my difference-in-differences estimates for the treatment effects rely on within firm-by-time variation for internal validity. For this reason, I create a subsample (*DiD sample*) that contains only observations of firms that form at least one pair that ever experienced travel time reductions due to the introduction of the high-speed railway.[30]

Table VI provides a summary for this sample. Statistical distributions of the outcome variables are similar between the DiD sample and the full sample. Although control pairs have longer distances, the two groups have similar travel times before the treatment. Due to data availability, I test the treatment effect on site visits using observations of SZSE-listed firms, which comprise roughly 70% of the sample. For trading profits, I use observations of both SSE-listed and SZSE-listed firms, while restricting the sample to SZSE-listed firms does not materially change the results. Moreover, the results remain similar if I exclude observations with extreme values of the outcome variables, so these results are not driven by a small number of influential outliers.

D.1 Main Results

Table VII reports the results of baseline regressions. In Column (1), the point estimate for the treatment effect on the number of site visits is close to 0.01 and statistically significant.[31] This is equivalent to a 25% increase relative to the unconditional average frequency of site visits, or 4.6% of a standard deviation. The effect on fund families' active portfolio weight,

[30] Results are qualitatively and quantitatively similar if I use the original full sample in the estimation, because pairs from other firms do not provide any within firm-by-time variation in the treatment status.

[31] Standard errors are two-way clustered at the fund family level and the firm industry level. There are 74 industry classes under the CSRC classification.

as shown in Column (2), is small and statistically insignificant. In Column (3), the dependent variable is replaced with investment profits measured in CNY millions. The point estimate is above 1.2, with t-statistics greater than 3.0. This implies that the travel time reductions lead to a CNY 1.2 million increase, on average, in trading profits (in 2006 CNY) during a 6-month period, or roughly 4.2% of a standard deviation. These results provide evidence that the travel time reductions have positive causal impact on information acquisition activities and investment performance. The difference between the effects on portfolio weights in Column (2) and trading profits in Column (3) is consistent with my finding that Chinese mutual funds trade frequently and realize investment profits mainly through intraperiod stock trades.

The introduction of high-speed rail lines should not directly affect fund managers' participation in remote meetings because these activities do not require physical travel between the fund family and firm. Such meetings include conference calls and meetings held at other locations. In Column (4), I conduct a placebo test using the number of remote meetings as the dependent variable. The coefficients are indistinguishable from zero, suggesting that the causal effects found here are indeed driven by reductions in travel-related information costs.

D.2 Dynamics

The predetermined nature of the high-speed rail line introductions alleviates the concern about endogenous event timing. To check whether the estimates are driven by pre-existing trends, I further examine the dynamics of the treatment effects by replacing the treatment dummy with a set of dummy variables: $Treatment(-2)$, $Treatment(-1)$, $Treatment(0)$, $Treatment(+1)$, $Treatment(+2)$, $Treatment(+3)$ and $Treatment(\geq +4)$. Specifically, $Treatment(-2)$ equals 1 if the observation is of a treated pair that experienced high-speed rail introduction two periods later, and other dummy variables are analogously defined relative to the event dates. These variables capture the "effects" of high-speed rail lines at different time periods relative to the introduction events.

Table VIII reports the estimation results. All estimated coefficients before the treatment events are statistically indistinguishable from zero. These results do not suggest that the (untestable) parallel-trend assumption underlying the difference-in-differences estimator is violated. Results in Columns (1) and (3) indicate that the effects of travel time reductions emerge quickly after the introduction of high-speed rail lines, and these effects persist over more than two years from these events. This indicates that fund managers indeed take advantage of the permanent shock to information costs to improve their investment by acquiring more private signals. Column (2) shows that there is a positive effect on active portfolio weights after the introduction of high-speed rail lines, although such an effect appears small and transient. Column (4) finds no effect in any period for remote meetings.

D.3 Intensity of the Treatment

Different fund family–firm pairs in the treated group experience treatment with different intensity, and larger reductions in travel times are likely to cause stronger effects. In Table IX, I examine how the treatment effects depend on the amount of travel time reductions. Based on Equation (3), I interact $Treatment$ with two dummy variables, *Large* and $Small$. The dummy variable *Large* equals 1 if the CRH network reduces the one-way travel time by more than one hour, and *Small* equals 1 if the travel time reduction is less than one hour. The treatment effects appear to be stronger for pairs that experience larger travel time reductions. The magnitude of the estimated effects for trading profits are considerably larger than the baseline specification, which suggests that fund managers gain more benefits of private information when they adjust information choices in response to larger reductions in information costs.

D.4 Cross-Sectional Heterogeneity

In Table X, I explore how fund family–firm pairs' different time-invariant characteristics give rise to heterogeneities in the treatment effects. To do so, I interact $Treatment$ with two groups of dummy variables.[32] I first divide all pairs into *Far* and *Near* groups, depending on whether the distance between the addresses of the mutual fund family and the firm headquarters is greater than 500 km. Column (1) shows that the effect on site visits is stronger for distant treated pairs, in which the magnitude is two-thirds larger than the baseline estimate. In Column (2), the point estimate for $Treatment \times Far$ is moderately larger than that of $Treatment \times Near$, although both are statistically significant. Consistent with Table IX, these results imply that the effects of travel time reductions are more important for long-distance travel.

Next, I divide all pairs into a *Manufacturing* group and an *Other-industry* group based on the firm's industry classification. In Columns (3)-(4), the coefficient estimates are similar to the baseline for both groups, but statistical significance is found mostly for the manufacturing group. The results of these tests suggest that the tangibility of firms' assets in manufacturing industries potentially improves the effectiveness of acquiring soft information on site.

D.5 External Validity

Under the identifying assumption, my causal estimates are internally valid. There are three caveats in interpreting these quantities and extrapolating them to more general settings. First, only firms that are both close to new railway stations and have medium distances to financial hubs are possibly affected by the introductions of high speed rail lines (as shown in Figure III). As a result, the treated pairs could be different from the universe of fund family–firm pairs, and the magnitudes of population average treatment effects may also be different from the estimated effects. Second, this study focuses on only one form of private information acquisition (i.e., site visits) and only one dimension of information costs (i.e., travel times). There are other forms of information acquisition with different dimensions of information costs. Finally, the empirical setting in this study is based on the Chinese market, so quantitative implications might be specific to this market.

[32] Since these characteristics are controlled by the fixed effects, the coefficients of interaction terms are identified.

V Conclusion

This study provides direct evidence on how much information costs causally affect investors' information choices and investment decisions. To overcome the empirical challenge of studying the information choice problem, I use data on Chinese mutual fund managers' visits to firm headquarters and their stock trades. I exploit the introduction of high-speed rail lines in China as a quasi-natural experiment to establish a causal link between travel times and mutual fund decisions.

If private information is exogenously endowed to fund managers, travel time reductions should not affect their decisions and performance. Results in this paper reject this null hypothesis. Controlling for firm-level shocks, I find that the introduction of high-speed rail lines leads to sizable increases in both the frequency of site visits and trading profits at the fund family–firm pair level. These findings suggest that fund managers actively trade off the costs and benefits in their acquisition of private information, thus providing evidence for a broad class of theories that feature investors' endogenous information choices.

This study also provides one step towards a better understanding of the active management industry. Although a large body of literature has been devoted to this industry, the various forms and associated costs of the key production process, namely information acquisition, remains largely unknown. My findings shed light on the importance of costs in the production of private information for mutual funds. A unified framework that incorporates richer demand and supply factors faced by asset managers, including money flows, management fees, information costs, and performance, is left to future research.

Figures

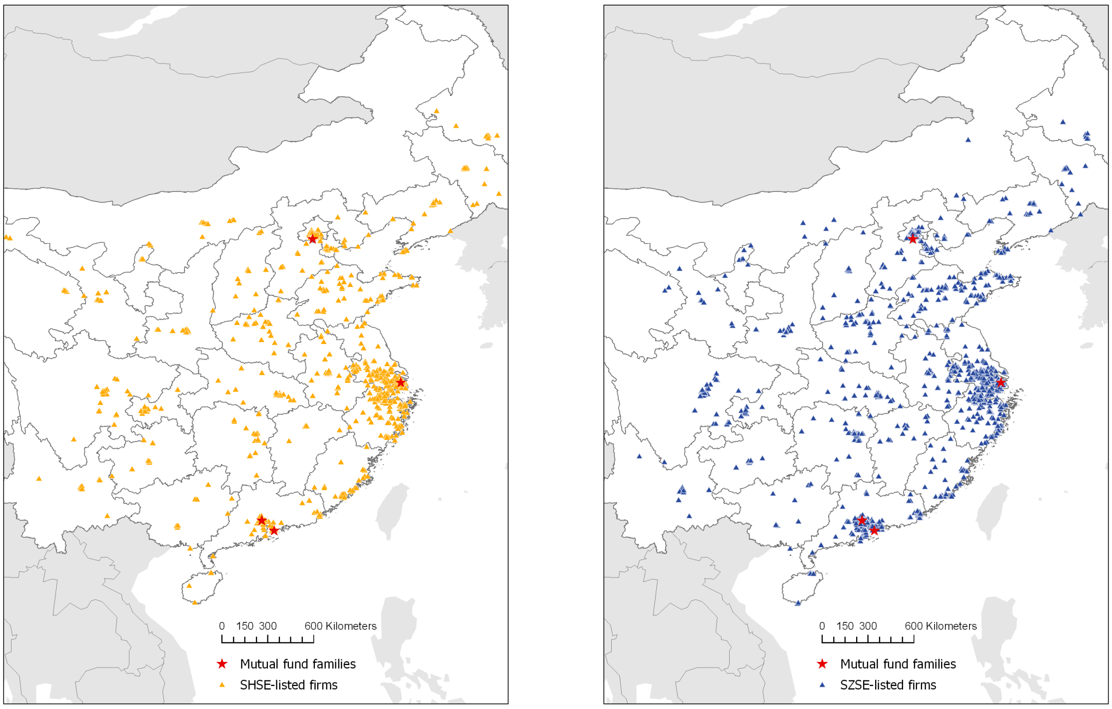


Figure I. Geographical distribution of sample firms and mutual fund families.
The left panel plots firm headquarters locations for all SSE-listed firms in the sample, and the right panel plots firm headquarters locations for SZSE-listed firms. In both panels, red stars denote mutual fund family office locations.

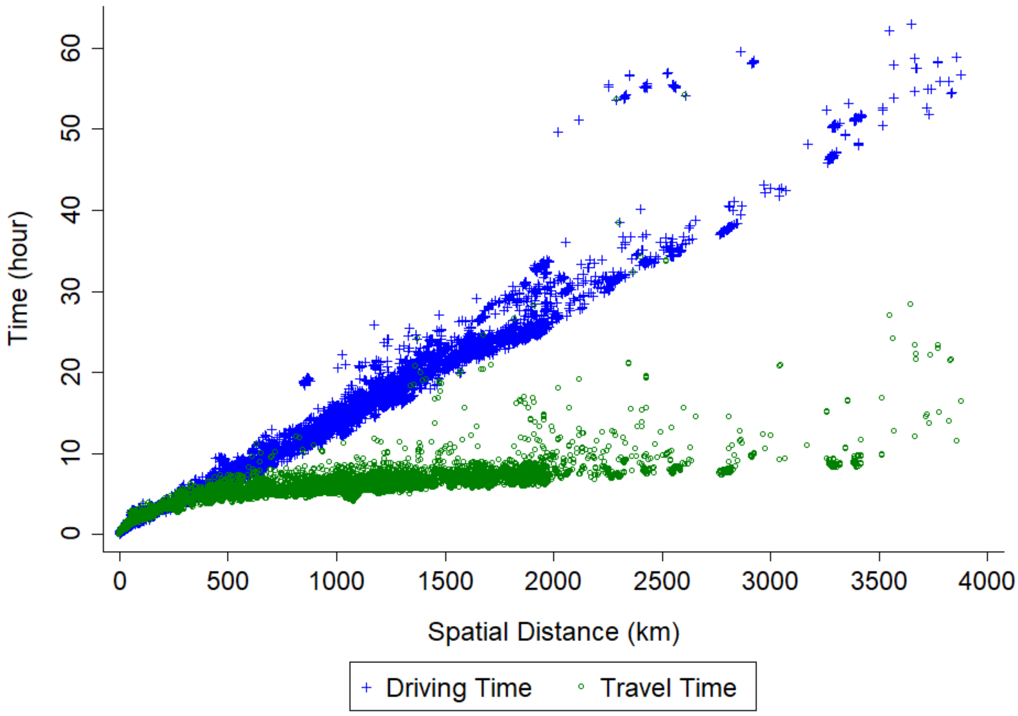
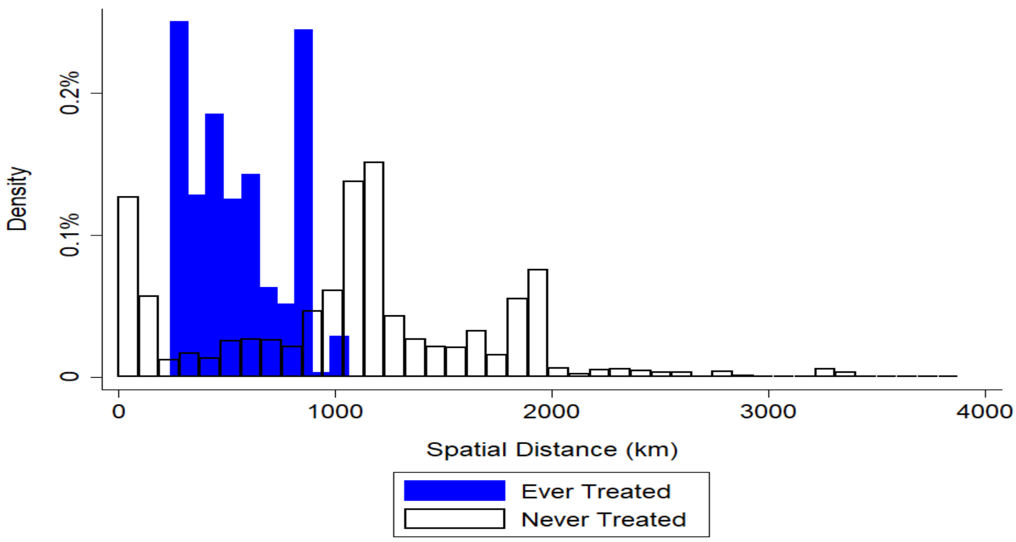


Figure II. Geographical distance and travel time.
This figure plots the geographical distance and travel time between locations of mutual fund families and firm headquarters. Each marker represents a pair composed of a fund family location and a firm location.



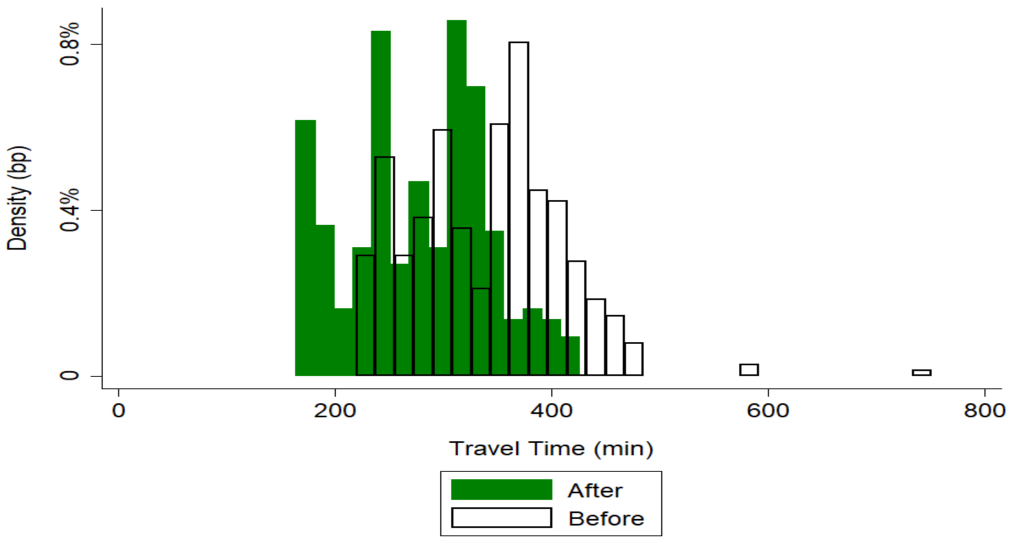


Figure III. Geographical distance and travel times of treated pairs.
This figure reports the distributions of geographical distance and travel time between fund families and firms in the treated pairs. The upper panel plots the distribution of distance (in kilometers) for pairs from the treated and the control groups. The lower panel plots the distribution of travel times (in minutes) before and after CRH connection events for pairs in the treatment group.

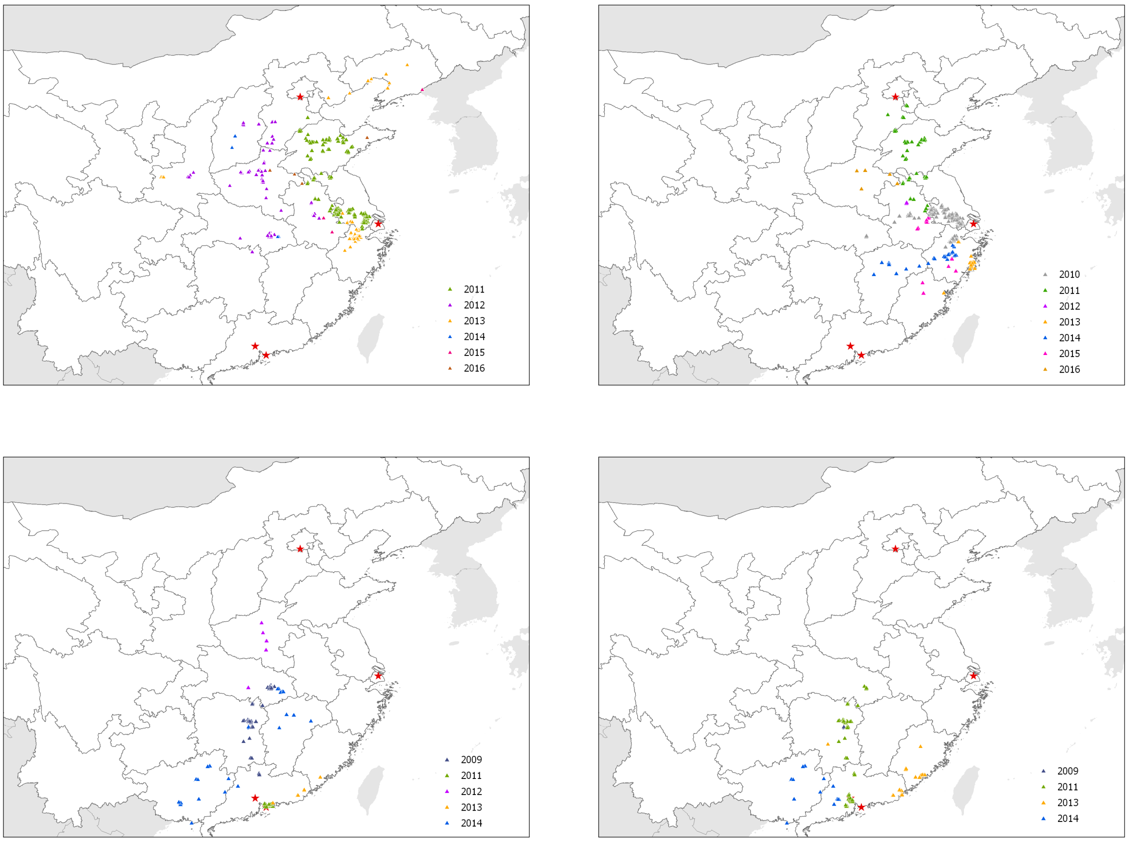


Figure IV. Geographical distribution of firms in treated pairs. Each of the four subfigures visualizes headquarters locations for firms in treated pairs associated with mutual fund families located in a corresponding financial hub. From upper left to lower right, the financial hub cities are: Beijing, Shanghai, Guangzhou, and Shenzhen. Different colors of firm markers indicate different treatment cohorts.

Table I. Sample Composition

This table summarizes the composition of the sample. Panel A reports the numbers of sample firms by their listing exchange (the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE)), the numbers of fund families by financial hub city where they are located (Beijing (BJ), Shanghai (SH), Guangzhou (GZ), and Shenzhen (SZ)), and the numbers of fund family–firm pairs. Panel B reports the industry distribution of sample firms.

Panel A: Number of Unique Firms, Fund Families and Pairs by Date

	# Firm			# Fund Family					# Pair
Period	SSE	SZSE	Total	BJ	SH	GZ	SZ	Total	Total
2008 Jun	713	596	1,309	13	30	3	12	58	75,922
2009 Dec	718	692	1,410	13	31	3	13	60	84,600
2011 Dec	770	1,216	1,986	13	34	3	14	64	127,104
2013 Dec	792	1,335	2,127	15	37	3	16	71	151,017
2015 Dec	914	1,541	2,455	28	48	3	20	99	243,045
2017 Dec	1,219	1,879	3,098	35	52	4	23	114	353,172

Panel B: Number of Sample Firms by Industry

CSRC Industry Classifcation Category	Code	# Firms	Fraction
Agriculture, forestry, animal husbandry and fishery	A	37	1.2%
Mining	B	61	1.9%
Manufacturing	C	2,067	66.0%
Electric power, heat, gas and water production	D	92	2.9%
Construction	E	87	2.8%
Wholesale and retail industry	F	151	4.8%
Transport, storage and postal service industry	G	90	2.9%
Accommodation and catering industry	H	9	0.3%
Information transmission, software and technology	I	234	7.5%
Real estate	K	98	3.1%
Leasing and commercial service	L	42	1.3%
Scientific research and technical service	M	44	1.4%
Water conservancy, environment and public facility	N	44	1.4%
Education	P	2	0.1%
Health and social work	Q	7	0.2%
Culture, sports and entertainment	R	49	1.6%
Diversified industries	S	17	0.5%
Total		3,131	100.0%

Table II. Summary Statistics

This table reports summary statistics of the full sample. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. $\#Visit$ and $\#Remote$ are the numbers of site visits and remote meetings. Panel A shows firm characteristics, and observations are aggregated to the firm–semi-year level. Firm market size, based on the number of tradable shares, is measured in CNY billions. $\#MFHolder$ and $MFHolding$ are the number of mutual fund families that hold the firm's stock and the fraction of market capitalization held by mutual funds, respectively. Panel B shows fund family characteristics, and observations are aggregated to the fund family–semi-year level. $\#StockHolding$ and $PortfolioValue$ are the number of stocks held and the total market value of stock holdings (in CNY billions), respectively. Panel C shows the main variables in the pair–semi-year panel. $PortfolioWeight$ is the weight of a stock in a fund family's portfolio. $Profit$ is calculated as $Profit_{t-1\to t} = Holding_t + Sell_{t-1\to t} - Buy_{t-1\to t} - Holding_{t-1}$, where Buy and $Sell$ are cumulative amounts of cash flows from intraperiod stock purchases and sales measured in CNY millions. The number of site visits and remote meetings are observed only for SZSE-listed firms.

	N	Mean	STD	p5	p25	p50	p75	p95
Panel A: Firm–Semi-year Level								
Size	42,576	9.0	42.9	0.6	1.5	3.1	6.4	24.8
# Visit	25,042	3.0	6.0	0.0	0.0	0.0	4.0	15.0
# Remote	25,042	0.2	2.3	0.0	0.0	0.0	0.0	0.0
# MF Holder	42,576	9.4	11.8	0.0	1.0	5.0	13.0	33.0
MF Holding	42,576	5.9%	10.1%	0.0%	0.1%	1.4%	7.0%	28.2%
Book-to-Market	41,123	1.0	1.5	0.2	0.3	0.6	1.1	2.8
ROE	42,555	4.3%	255.4%	-5.0%	2.0%	4.8%	8.9%	17.7%
Stock Return	42,314	8.2%	44.9%	-36.9%	-18.7%	-2.0%	22.3%	88.4%
Panel B: Fund Family–Semi-year Level								
# Visit	1,529	49.5	41.7	3.0	18.0	39.0	72.0	126.0
# Remote	1,529	3.9	4.3	0.0	1.0	3.0	6.0	12.0
# Stock Holding	1,529	261.1	217.5	30.0	119.0	202.0	333.0	731.0
Portfolio Value	1,529	13.9	16.9	0.1	1.8	7.5	20.1	49.1

Panel C: Pair–Semi-year Level

	All Observations				Nonzero-Valued Observations			
	N	Mean	STD		N	p5	p50	p95
# Visit	2,017,547	0.04	0.21		68,698	1.0	1.0	2.0
# Remote	2,017,547	0.0	0.1		5,496	1.0	1.0	2.0
Portfolio Weight	3,347,795	0.0%	0.3%		399,199	0.0%	0.1%	1.7%
Profit	2,987,068	0.2	33.6		602,319	-55.9	0.0	62.1

Table III. Site Visits and Investment Activities

This table reports regression estimates for the relation between site visits and mutual fund investment activities. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. In Columns (1) and (2), the dependent variable # Visit$_{t-1\to t}$ is the number of site visits during a 6-month period, and the independent variable Active Weight$_{t-1}$ is measured in percentage points. In Column (3), the dependent variable Active Weight$_t$ is measured in basis points. In Column (4), the dependent variable Hold$_t$ is a dummy variable that equals one if the fund family holds the firm's stock at the end of period t. In Column (5), the dependent variable Trade$_{t-1\to t}$ is a dummy variable that equals one if the fund family trades (buys or sells) the firm's stock during a 6-month period. Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	# Visit$_{t-1\to t}$		Active Weight$_t$	Hold$_t$	Trade$_{t-1\to t}$
	(1)	(2)	(3)	(4)	(5)
Hold$_{t-1}$	0.017***				
	(13.271)				
Active Weight$_{t-1}$		0.017***			
		(8.034)			
# Visit$_{t-1\to t}$			5.338***	0.084***	0.059***
			(10.989)	(27.661)	(20.778)
Pair FEs	Yes	Yes	Yes	Yes	Yes
Firm × Time FEs	Yes	Yes	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes	Yes	Yes
R^2	0.249	0.250	0.330	0.390	0.403
Observations	1,803,246	1,803,246	2,002,827	2,002,827	2,002,827

Table IV. Site Visits, Intraperiod Trades and Investment Profits

This table presents the empirical relationship among mutual fund site visits, intraperiod stock trades, and investment profits. Investment profits is measured as $Profit_{t-1\to t} = Holding_t + Sell_{t-1\to t} - Buy_{t-1\to t} - Holding_{t-1}$, where *Buy* and *Sell* are cumulative amounts of cash flows from intraperiod stock purchases and sales, measured in CNY millions. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. In Panel A, portfolios are formed based on (a) whether the firm is visited during the period and (b) whether the firm's stock is traded during the period. Arithmetic means are first calculated within each portfolio and then calculated over periods. $Profit(byFirm)\%$ and $Profit(byFamily)\%$ are profit scaled by lagged firm market capitalization and by lagged fund family's total value of stock holding, respectively, in basis points. Panel B reports the results of regressing investment profits on Trade$_{t-1\to t}$, a dummy variable that equals one if the fund family trades (buys or sells) the firm's stock, and the number of site visits during a 6-month period. Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

Panel A: Average Investment Profit						
	Visit	Trade	Profit	Profit(byFirm)%	Profit(byFamily)%	Observations
(1)	Y	N	1.46	10.13	1.94	49,388
(2)	N	N	-0.09	0.43	0.40	1,634,781
(3)	Y	Y	4.87	25.42	6.80	14,989
(4)	N	Y	2.86	11.03	6.57	120,941

Panel B: Regression Estimates			
	Profit$_{t-1\to t}$		
	(1)	(2)	(3)
Trade$_{t-1\to t}$	2.141^{***}		2.078^{***}
	(4.789)		(4.717)
# Visit$_{t-1\to t}$		1.369^{***}	1.246^{***}
		(3.376)	(3.127)
Pair FEs	Yes	Yes	Yes
Firm × Time FEs	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes
R^2	0.109	0.109	0.109
Observations	1,771,953	1,771,953	1,771,953

Table V. Proximity, Site Visits, and Investment

This table reports regression estimates for the effects of pairwise travel time (in hour) and distance (in thousand kilometers) on mutual fund site visits and investment decisions. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. In Columns (1)-(3), the dependent variable is the number of site visits. In Columns (4)-(6), the dependent variable is active portfolio weight (in basis points). In Columns (7)-(9), the dependent variable is a dummy variable that equals one if the fund family trades (buys or sells) the firm's stock. In Columns (10)-(12), the dependent variable is trading profits (in CNY millions). Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

Panel A: Site Visits and Portfolio Active Weights

	Visit			Active Weight		
	(1)	(2)	(3)	(4)	(5)	(6)
Travel Time	-0.007^{***}		-0.005^{***}	-0.101^{***}		-0.066^{**}
	(-8.368)		(-3.219)	(-5.578)		(-2.427)
Distance		-0.022^{***}	-0.008^{*}		-0.334^{***}	-0.135
		(-9.095)	(-1.856)		(-5.078)	(-1.276)
Firm × Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.147	0.147	0.147	0.278	0.278	0.278
Observations		2,017,565			3,347,813	

Panel B: Stock Trades and Trading Profits

	Trade			Profit		
	(7)	(8)	(9)	(10)	(11)	(12)
Travel Time	-0.001^{***}		-0.001^{**}	-0.074^{**}		-0.073^{*}
	(-4.456)		(-2.446)	(-2.428)		(-1.670)
Distance		-0.004^{***}	-0.001		-0.229^{**}	-0.008
		(-4.370)	(-0.421)		(-2.341)	(-0.061)
Firm × Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.294	0.294	0.294	0.071	0.071	0.071
Observations		3,347,813			2,987,086	

Table VI. Summary of the Difference-in-Differences Sample

This table summarizes the difference-in-differences sample. Panel A reports the distributions of variables for pairs in the treated and control groups. $PortfolioWeight$ is the weight of a stock in a fund family portfolio, measured in basis points. $Profit$ is trading profit, calculated as $Profit_{t-1\to t} = Holding_t + Sell_{t-1\to t} - Buy_{t-1\to t} - Holding_{t-1}$, where Buy and $Sell$ are cumulative values of mutual fund intraperiod stock trades measured in CNY millions. $TravelTime$ for Ever Treated pairs is the travel time before the introduction of high-speed rail. Panel B reports the numbers of pairs that experience travel time reductions after each group of high-speed railway introduction events.

Panel A: Summary Statistics						
	Ever Treated			Never Treated		
	N	Mean	STD	N	Mean	STD
# Visit	72,141	0.04	0.21	190,859	0.03	0.19
# Remote	72,141	0.00	0.05	190,859	0.00	0.05
Portfolio Weight	127,483	0.03%	0.27%	293,318	0.03%	0.26%
Profit	114,240	0.17	28.63	265,648	0.06	31.94
Distance (km)	127,483	502.9	221.2	293,318	1,093.2	379.3
Travel Time (min)	127,483	327.0	67.0	293,318	366.0	78.3

Panel B: Number of Treated Pairs by Event Year	
Event Year	# Pairs in Treatment Group
2009	181
2010	5,353
2011	4,952
2012	1,738
2013	257
2014	421
2015	33
2016	255
Total	13,190

Table VII. Difference-in-Differences: Main Regressions

This table reports results from estimating regression

$$Outcome_{i,f,t} = \beta \times Treatment_{i,f,t} + \alpha_i \times \alpha_f + \alpha_f \times \delta_t + \alpha_i \times \delta_t + \varepsilon_{i,f,t},$$

where $Treatment_{i,f,t}$ is a dummy variable that equals 1 if high-speed rail that reduces travel time between the office locations of fund family i and firm f is in service during period t. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. Dependent variables are the number of site visits, active portfolio weights, and trading profits. Column (4) reports a placebo test where the dependent variable is the number of fund managers' participation in private meetings with the firm that do not occur on site (either conference calls or held at different locations). Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	Visit	Active Weight	Profit	Remote
	(1)	(2)	(3)	(4)
Treatment	0.009***	0.323	1.268***	0.000
	(3.422)	(1.126)	(3.082)	(0.170)
Pair FEs	Yes	Yes	Yes	Yes
Firm × Time FEs	Yes	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes	Yes
R^2	0.228	0.354	0.131	0.261
Observations	263,000	420,801	374,025	263,000

Table VIII. Dynamics of Treatment Effects

This table reports the dynamics of estimated treatment effects. All dependent variables are defined as in Table VII. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. $Treatment(-2)$ is a dummy variable that equals 1 if the observation is from a treated pair that experiences a travel time reduction two periods later. $Treatment(-1)$, $Treatment(0)$, $Treatment(1)$, $Treatment(2)$, $Treatment(3)$, and $Treatment(4+)$ are defined analogously. Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	# Visit	Active Weight	Profit	# Remote
	(1)	(2)	(3)	(4)
Treatment (-2)	-0.002	0.258	-0.919	0.000
	(-0.389)	(0.485)	(-0.503)	(0.421)
Treatment (-1)	0.004	0.530	0.623	0.001
	(0.507)	(0.795)	(0.880)	(0.780)
Treatment (0)	0.004	1.224	0.496	0.004
	(0.431)	(1.610)	(0.357)	(1.641)
Treatment (+1)	0.020**	1.116*	1.547***	0.000
	(2.645)	(1.803)	(3.007)	(0.406)
Treatment (+2)	0.014*	0.701	1.622**	0.000
	(1.759)	(1.347)	(2.147)	(0.504)
Treatment (+3)	0.023***	1.129**	0.992**	0.001
	(3.211)	(2.275)	(2.205)	(1.076)
Treatment ($\geq$+4)	0.008***	0.566	1.335**	0.001
	(2.764)	(1.607)	(2.541)	(1.178)
Pair FEs	Yes	Yes	Yes	Yes
Firm $\times$ Time FEs	Yes	Yes	Yes	Yes
Fund Family $\times$ Time FEs	Yes	Yes	Yes	Yes
R^2	0.226	0.354	0.131	0.262
Observations	263,000	420,801	374,025	263,000

Table IX. Intensity of the Treatment

This table reports estimated effects in response to different intensity of travel time reductions. All dependent variables are defined as in Table VII. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. *Large* is a dummy variable that equals 1 if the introduction of high-speed rail lines reduces travel time between office locations in a pair by at least 60 minutes in a one-way trip, and *Small* is a dummy variable that equals 1 if the travel time reduction is less than 60 minutes. Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	Visit	Profit
	(1)	(2)
Treatment Large	0.010**	1.772***
	(2.601)	(3.801)
Treatment Small	0.007	0.561
	(1.436)	(1.478)
Pair FEs	Yes	Yes
Firm × Time FEs	Yes	Yes
Fund Family × Time FEs	Yes	Yes
R^2	0.228	0.131
Observations	263,000	374,025

Table X. Cross-Sectional Heterogeneity of Treatment Effect

This table reports estimated treatment effects from different pairs in the treated group. All dependent variables are defined as in Table VII. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. *Far* (*Near*) is a dummy variable that equals 1 if the geographical distance between the two locations in a pair is larger (smaller) than 500 km. *Manufacturing* (*OtherIndustry*) is a dummy variable that equals 1 if the firm of a pair belongs (does not belong) to manufacturing industries. Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	Distance		Firm Industry	
	# Visit	Profit	# Visit	Profit
	(1)	(2)	(3)	(4)
Treatment × Far	0.015***	1.326**		
	(3.535)	(2.077)		
Treatment × Near	0.002	1.109*		
	(0.354)	(2.001)		
Treatment × Manufacturing			0.009***	1.293**
			(2.919)	(2.065)
Treatment × OtherIndustry			0.007	1.064*
			(1.606)	(1.799)
Pair FEs	Yes	Yes	Yes	Yes
Firm × Time FEs	Yes	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes	Yes
R^2	0.228	0.129	0.228	0.129
Observations	263,000	374,025	263,000	374,025

References

Bernstein, S., X. Giroud, and R. R. Townsend (2016). The impact of venture capital monitoring. *The Journal of Finance 71*(4), 1591–1622.

Bowen, R. M., S. Dutta, S. Tang, and P. Zhu (2018). Inside the "black box" of private in-house meetings. *Review of Accounting Studies 23*(2), 487–527.

Bullock, R., J. Sondhi, and P. Amos (2009). Tracks from the past, transport for the future-China's railway industry 1990-2008 and its future plans and possibilities. *Beijing: the World Bank Transport Coordinator*.

Chen, H., L. Cohen, U. G. Gurun, D. Lou, and C. J. Malloy (2017). IQ from IP: Simplifying search in portfolio choice.

Chen, H., Y. Qu, T. Shen, and Q. Wang (2019). The geography of information acquisition. In *SFS Cavalcade North America 2019*.

Chen, Y., B. Kelly, and W. Wu (2018). Sophisticated investors and market efficiency: Evidence from a natural experiment. Technical report, National Bureau of Economic Research.

Cheng, Q., F. Du, X. Wang, and Y. Wang (2016). Seeing is believing: analysts' corporate site visits. *Review of Accounting Studies 21*(4), 1245–1286.

Cheng, Q., B. Y. Wang, and X. Wang (2017). Do corporate site visits impact stock prices? *Contemporary Accounting Research*.

Cohen, L., A. Frazzini, and C. Malloy (2008). The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy 116*(5), 951–979.

Coval, J. D. and T. J. Moskowitz (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance 54*(6), 2045–2073.

Coval, J. D. and T. J. Moskowitz (2001). The geography of investment: Informed trading and asset prices. *Journal of political Economy 109*(4), 811–841.

Crane, A. D., K. Crotty, and T. Umar (2018). Do hedge funds profit from public information?

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

Da, Z., U. G. Gurun, B. Li, and M. Warachka (2020). Investment in a smaller world: The implications of air travel for investors and firms. *Management Science*.

Dong, R., R. Fisman, Y. Wang, and N. Xu (2019). Air pollution, affect, and forecasting bias: Evidence from chinese financial analysts. *Journal of Financial Economics*.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2012). Investor information demand: Evidence from google searches around earnings announcements. *Journal of Accounting research 50*(4), 1001–1040.

Ellis, J. A., L. Madureira, and S. Underwood (2019). The causal effects of proximity on investment: Evidence from flight introductions. *Journal of Financial and Quantitative Analysis*.

Gallagher, E. A., L. D. Schmidt, A. Timmermann, and R. Wermers (2018). Investor information acquisition and money market fund risk rebalancing during the 2011–2012 eurozone crisis. *The Review of Financial Studies*.

Gao, M. and J. Huang (2016). Capitalizing on capitol hill: Informed trading by hedge fund managers. *Journal of Financial Economics 121*(3), 521–545.

Gargano, A., A. G. Rossi, and R. Wermers (2017). The freedom of information act and the race toward information acquisition. *The Review of Financial Studies 30*(6), 2179–2228.

Gârleanu, N. and L. H. Pedersen (2018). Efficiently inefficient markets for assets and asset management. *The Journal of Finance 73*(4), 1663–1712.

Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *The Quarterly Journal of Economics 128*(2), 861–915.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American economic review 70*(3), 393–408.

Han, B., D. Kong, and S. Liu (2018). Do analysts gain an informational advantage by visiting listed companies? *Contemporary Accounting Research*.

Holmström, B. and J. Tirole (1993). Market liquidity and performance monitoring. *Journal of Political Economy 101*(4), 678–709.

Hong, H., J. D. Kubik, and J. C. Stein (2005). Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers. *The Journal of Finance 60*(6), 2801–2824.

Iliev, P. and M. Lowry (2014). Are mutual funds active voters? *The Review of Financial Studies 28*(2), 446–485.

Irvine, P., M. Lipson, and A. Puckett (2006). Tipping. *The Review of Financial Studies 20*(3), 741–768.

Jiang, F. and K. A. Kim (2015). Corporate governance in china: A modern perspective.

Massa, M. and Z. Rehman (2008). Information flows within financial conglomerates: Evidence from the banks–mutual funds relation. *Journal of Financial Economics 89*(2), 288–306.

Matvos, G. and M. Ostrovsky (2010). Heterogeneity and peer effects in mutual fund proxy voting. *Journal of Financial Economics 98*(1), 90–112.

Pool, V. K., N. Stoffman, and S. E. Yonker (2015). The people in your neighborhood: Social interactions and mutual fund portfolios. *The Journal of Finance 70*(6), 2679–2732.

Puckett, A. and X. Yan (2011). The interim trading skills of institutional investors. *The Journal of Finance 66*(2), 601–633.

Subrahmanyam, A. and S. Titman (1999). The going-public decision and the development of financial markets. *The Journal of Finance 54*(3), 1045–1082.

Sun, X., Y. Zhang, and S. Wandelt (2017). Air transport versus high-speed rail: an overview and research agenda. *Journal of Advanced Transportation 2017*.

Appendix for

“Costly Information Acquisition and Investment Decisions: Quasi-Experimental Evidence”

This Appendix presents supplemental materials to the empirical analysis in “Costly Information Acquisition and Investment Decisions: Quasi-Experimental Evidence”. Section A tabulates supplementary results, including a number of robustness checks for the difference-in-differences estimates. Section B presents figures related to the empirical setting. Section C and Section D describe the computation of pairwise travel times and hand collection of mutual fund company visits data, respectively. Section E is a collection of additional discussions.

A Supplementary Results

Table A.1. Summary of Mutual Fund Intraperiod Stock Trading

This table summarizes mutual fund stock trades over the semi-year horizon. Hold_{t-1} and Hold_t indicate whether a fund family holds a firm’s stock at the ends of the previous period and the current period, respectively. $\text{Buy}_{t-1\to t}$ and $\text{Sell}_{t-1\to t}$ indicate whether a fund family purchases and sells a firms’s stock during the current period, respectively. Stock purchases and sales are determined by whether nonzero trading cash flows are reported.

Hold_{t-1}	$\text{Buy}_{t-1\to t}$	$\text{Sell}_{t-1\to t}$	Hold_t	Percentage
N	N	N	N	91.2%
N	Y	N	Y	0.8%
N	Y	Y	Y	1.0%
N	Y	Y	N	1.8%
Y	N	N	Y	3.5%
Y	Y	N	Y	0.5%
Y	N	Y	Y	0.5%
Y	N	Y	N	0.7%

Table A.2. Robustness: Exclusion of Fund Families with Multiple Office Locations

This table reports results from re-estimating the regressions in Table VII while excluding pairs that belong to fund families with more than one office locations. Each regression estimates

$$Outcome_{i,f,t} = \beta \times Treatment_{i,f,t} + \alpha_i \times \alpha_f + \alpha_f \times \delta_t + \alpha_i \times \delta_t + \varepsilon_{i,f,t},$$

where $Treatment_{i,f,t}$ is a dummy variable that equals 1 if high-speed rail that reduces travel time between the office locations of fund family i and firm f is in service during period t. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. Dependent variables are the number of site visits, active portfolio weights, and trading profits. Column (4) reports a placebo test where the dependent variable is the number of mutual fund participation in private meetings with the firm that do not occur on site (either conference calls or held at different locations). Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	Visit	Active Weight	Profit	Remote
	(1)	(2)	(3)	(4)
Treatment	0.009**	0.368	1.390***	0.000
	(2.657)	(1.080)	(3.246)	(0.035)
Pair FEs	Yes	Yes	Yes	Yes
Firm × Time FEs	Yes	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes	Yes
R^2	0.221	0.366	0.124	0.256
Observations	229,454	366,102	323,633	229,454

Table A.3. Robustness: Winsorization of Observations with Large Values

This table reports results from re-estimating the regressions in Table VII while winsorizing the outcome variables. Each regression estimates

$$Outcome_{i,f,t} = \beta \times Treatment_{i,f,t} + \alpha_i \times \alpha_f + \alpha_f \times \delta_t + \alpha_i \times \delta_t + \varepsilon_{i,f,t},$$

where $Treatment_{i,f,t}$ is a dummy variable that equals 1 if high-speed rail that reduces travel time between the office locations of fund family i and firm f is in service during period t. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. Dependent variables are the number of site visits, active portfolio weights, and trading profits. Column (4) reports a placebo test where the dependent variable is the number of mutual fund participation in private meetings with the firm that do not occur on site (either conference calls or held at different locations). In Columns (1) and (4), observations with values greater than one are replaced with one. In Columns (2)-(3), the dependent variables are winsorized at the 0.5% and 99.5% levels (approximately equivalent to 2.5% and 97.5% levels among nonzero-valued observations for trading profits). Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	Visit	Active Weight	Profit	Remote
	(1)	(2)	(3)	(4)
Treatment	0.008***	0.226	0.723***	0.000
	(3.564)	(1.016)	(2.806)	(0.462)
Pair FEs	Yes	Yes	Yes	Yes
Firm $\times$ Time FEs	Yes	Yes	Yes	Yes
Fund Family $\times$ Time FEs	Yes	Yes	Yes	Yes
R^2	0.221	0.366	0.124	0.248
Observations	263,000	420,801	374,025	263,000

Table A.4. Robustness: Treatment Window

This table reports results from re-estimating the regressions in Table VII while requiring each treated pair to exist for at least 6 semi-year periods before and after the treatment events. Each regression estimates

$$Outcome_{i,f,t} = \beta \times Treatment_{i,f,t} + \alpha_i \times \alpha_f + \alpha_f \times \delta_t + \alpha_i \times \delta_t + \varepsilon_{i,f,t},$$

where $Treatment_{i,f,t}$ is a dummy variable that equals 1 if high-speed rail that reduces travel time between the office locations of fund family i and firm f is in service during period t. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. Dependent variables are the number of site visits, active portfolio weights, and trading profits. Column (4) reports a placebo test where the dependent variable is the number of mutual fund participation in private meetings with the firm that do not occur on site (either conference calls or held at different locations). Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	Visit	Active Weight	Profit	Remote
	(1)	(2)	(3)	(4)
Treatment	0.010**	0.637*	1.406**	0.000
	(2.149)	(1.725)	(2.606)	(0.135)
Pair FEs	Yes	Yes	Yes	Yes
Firm × Time FEs	Yes	Yes	Yes	Yes
Fund Family × Time FEs	Yes	Yes	Yes	Yes
R^2	0.232	0.370	0.142	0.288
Observations	204,597	323,587	290,699	204,597

Table A.5. Robustness: Scaled Trading Profits

This table reports results from re-estimating the regressions in Columns (5)-(6) of Table VII. Each regression estimates

$$Profit_{i,f,t} = \beta \times Treatment_{i,f,t} + \alpha_i \times \alpha_f + \alpha_f \times \delta_t + \alpha_i \times \delta_t + \varepsilon_{i,f,t},$$

where $Treatment_{i,f,t}$ is a dummy variable that equals 1 if high-speed rail that reduces travel time between the office locations of fund family i and firm f is in service during period t. Observations are at the pair–semi-year level, where each pair is a fund family and a firm. Dependent variables are scaled trading profits in basis points. In Column (1), trading profits are scaled by lagged firm market capitalization. In Column (2), trading profits are scaled by lagged fund family total market value of stock holdings. Standard errors are two-way clustered at the fund family level and the firm's CSRC industry class level. t-statistics are reported in parentheses. *, **, *** represent 10%, 5%, and 1% levels of significance.

	Profit Scaled by Firm	Profit Scaled by Fund Family
	(1)	(2)
Treatment	1.375**	0.477**
	(2.417)	(2.047)
Pair FEs	Yes	Yes
Firm × Time FEs	Yes	Yes
Fund Family × Time FEs	Yes	Yes
R^2	0.095	0.372
Observations	374,025	374,025

B Figures

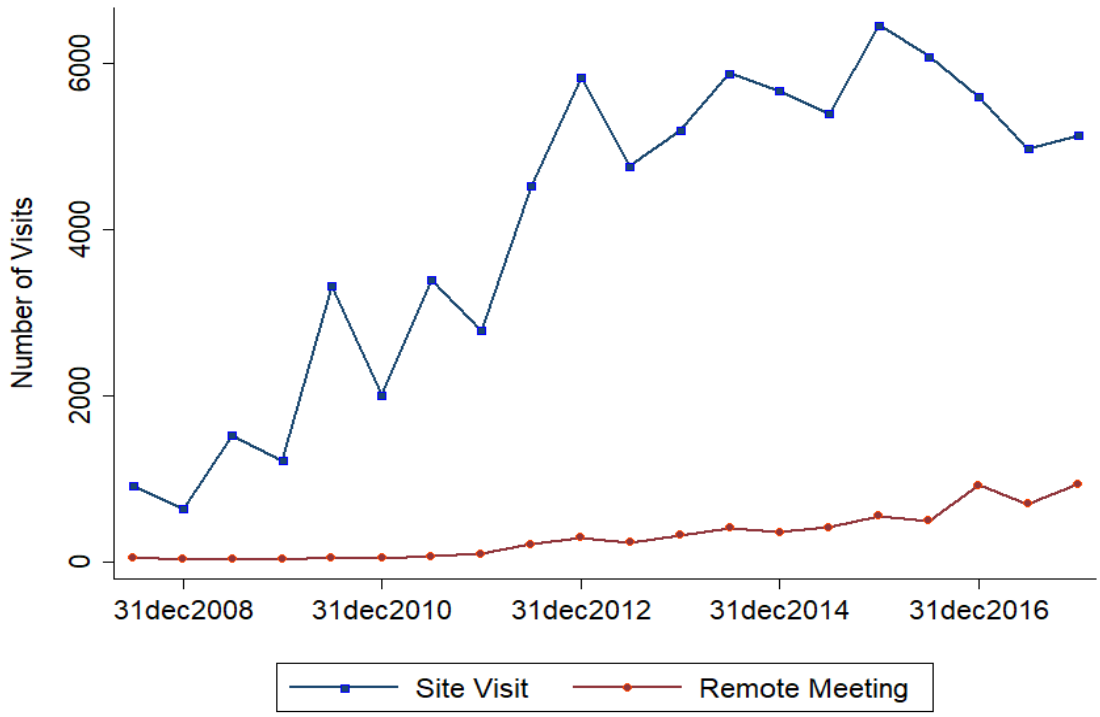
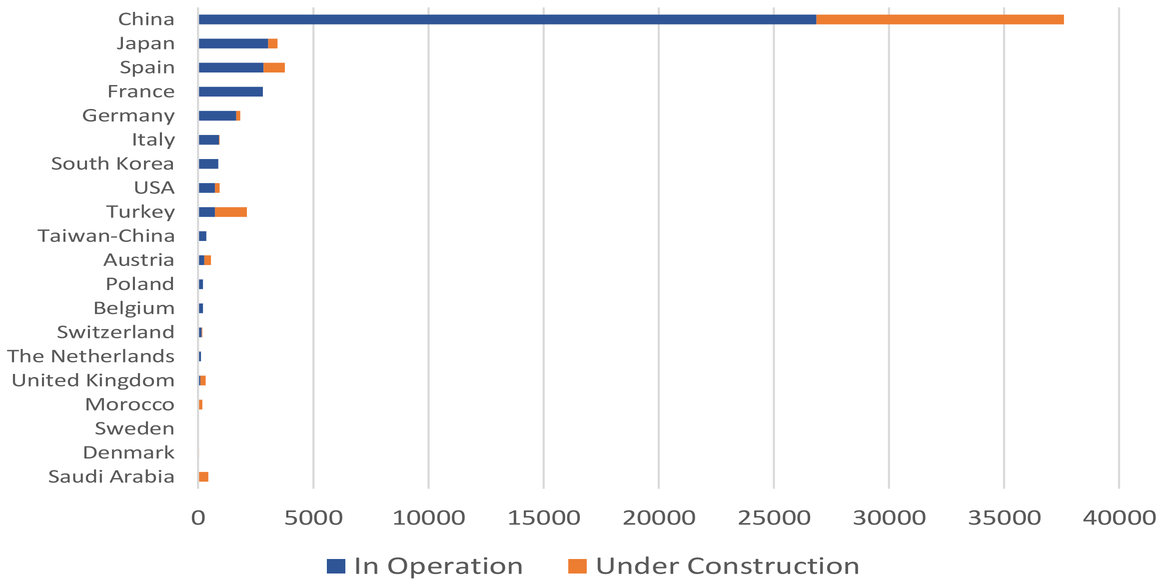


Figure A.1. Total number of mutual fund visits to SZSE-listed firms. This figure plots the time series of total mutual fund visits to firms listed on the Shenzhen Stock Exchange during each 6-month period. Site visits are defined as private meetings held at the firm's headquarters office. Remote meetings include conference calls and physical meetings at locations other than the firm's headquarters.

(a) High-speed rail length worldwide, 2017



(b) CRH annual growth rate, 2007-2017

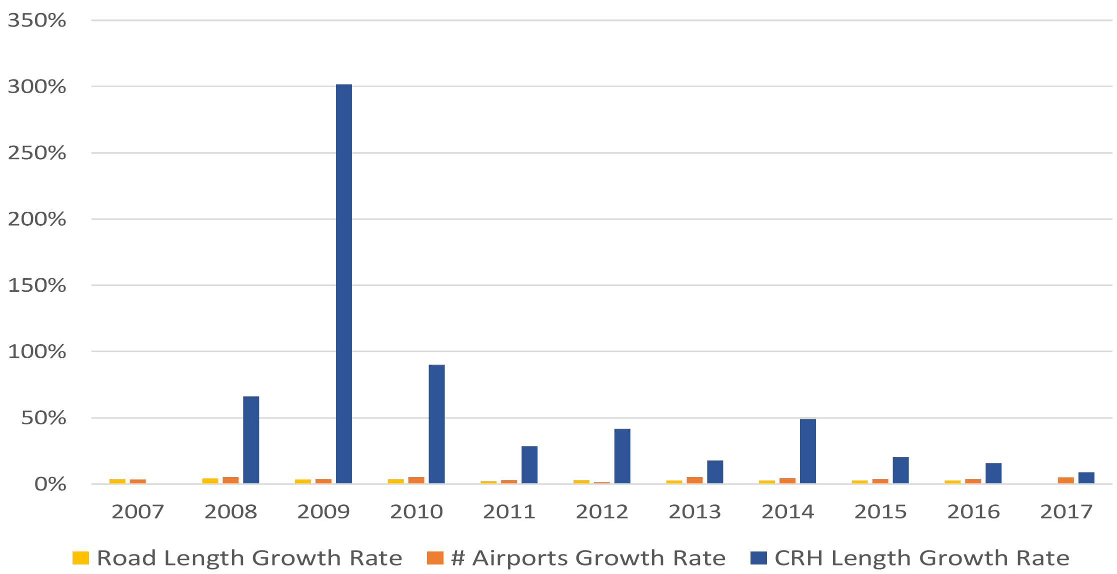


Figure A.2. Development of high-speed rail network in China and the world. This figure illustrates growth in the length of China Rail High-Speed (CRH) network. Panel (a) compares the CRH length with high-speed rail lengths in other countries, and Panel (b) compares CRH annual growth rate with development of roads and airports in China. Data sources: Yearbook of China Transportation & Communications, Civil Aviation Administration of China, National Railway Administration of the People's Republic of China, and the International Union of Railways (UIC), the World Railway Organization.

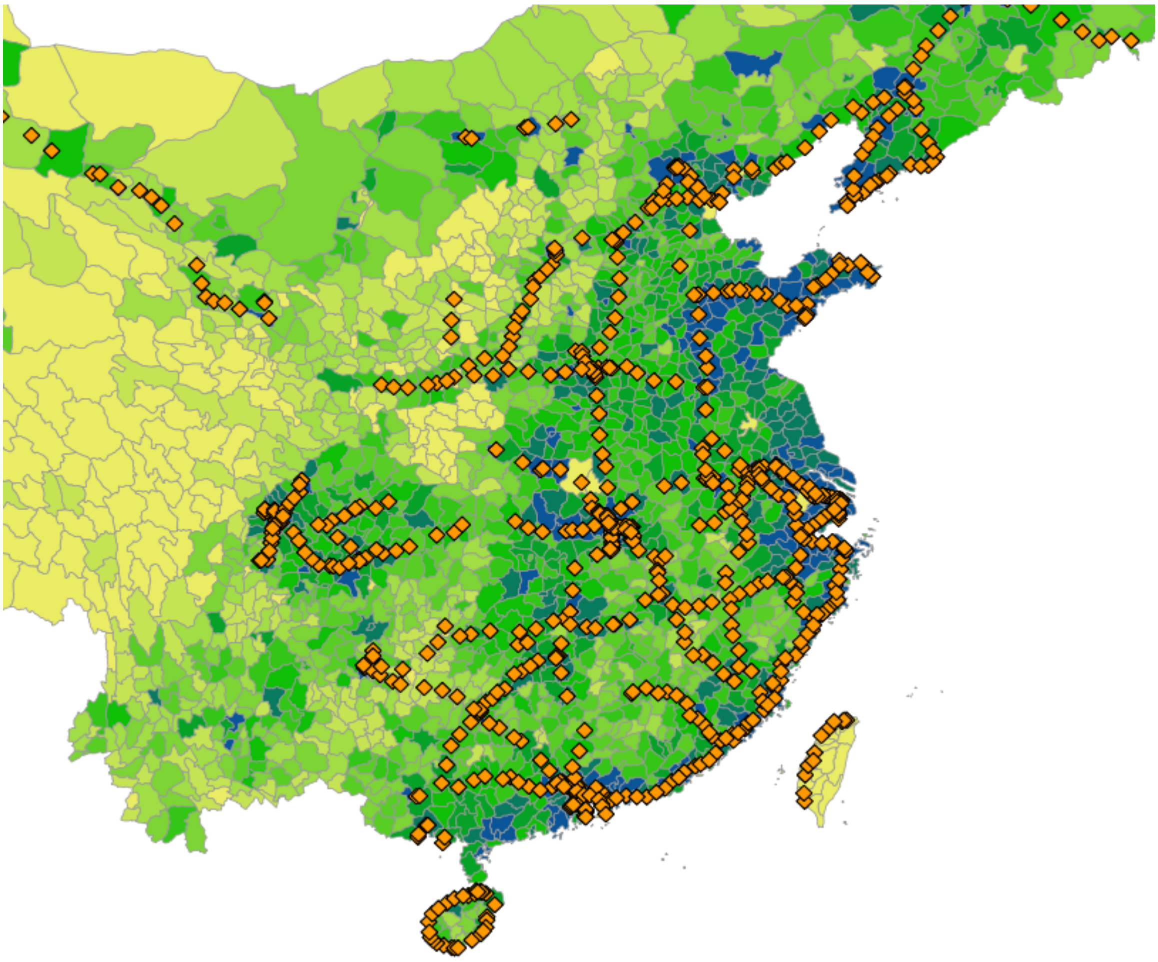


Figure A.3. CRH station locations at the end of 2016. Different colors in the background visualize county-level GDP measured in 2000. Data source: Harvard ChinaMap project.

C Travel Time Computation

To compute travel time estimates for a large number of origins and destinations, I use Web API services from two commercial mapping and navigation service providers in China: AMap and BaiduMap. Neither of these two strictly dominates the other when I perform this task, so I combine them for better estimation performance. In general, AMap does a better job of converting a string of address to accurate latitude and longitude coordinates, and it is superior in computing ground public transport time. The advantage of BaiduMap is that it includes air transport as a travel option. Given these facts, I use BaiduMap only for generating data for the flight segment (i.e., airport names and flight times), and I use AMap to perform the rest of the computation.

For the train-based plan, I force the API to prioritize trains. If there is at least one plan available, travel time is computed as the total time spent during these four trip segments:

1. Driving time from the origin to the departure railway station.
2. Time spent on the train.
3. Driving time from the arrival railway station to the destination.
4. The unobservable time spent in railway stations, which is assumed to be 60 minutes.

For the flight-based plan, I force the API to prioritize air transport. If there is at least one plan available, the travel time equals the sum of:

1. Driving time from the origin to the departure airport.
2. The time length of the flight;
3. Driving time from the arrival airport to the destination;
4. The unobservable time spent in airports, including take-off, landing, and potential delays, which is assumed to be 120 minutes in total.[33] The only exception is that, for flights between Beijing and Shanghai, I assume this time to be 60 minutes, because the introduction of Beijing-Shanghai Air Express service in 2007 greatly expedited the boarding process.[34]

Within each of these three travel plans, whenever there are multiple feasible options for each segment, I choose the combination that gives the shortest total travel time.

[33] This is consistent with the assumption in Sun, Zhang, and Wandelt (2017).

[34] The service included express check-in, security check, boarding gate and baggage claim services dedicated to the service at the two airports. For more information, see http://www.chinadaily.com.cn/china/2007-08/06/content_5448686.htm.

These estimates are computed around 10:00 am (Beijing time) on a representative business day in year 2018 to better reflect travel conditions faced by financial professionals. Since the amount of computation is large, I simultaneously execute 100 programs to ensure they are finished within 10 minutes, so the time estimates do not suffer from systematic incomparability in intraday traffic conditions.[35]

A potential concern is that these estimates might not reflect historical traffic conditions. It is difficult to calculate historical travel time because the navigation applications only provide travel plans based on currently available transportation, which gives rise to a data limitation. However, as shown in Figure A.2, there has been limited change in car or air transport during the sample period. Since most sample firms are located in reasonably accessible areas, the effect of changes in transports other than CRH should not materially bias my results. Moreover, if there are omitted overall improvements in other transportation, my travel time estimation would lead to conservative estimates for CRH-induced travel time shocks and treatment effects.

[35] Sequentially executing the program takes more than 10 hours, and different travel times would be computed under different traffic conditions depending on the time of day.

D Site Visit Data Collection

The China Securities Regulatory Commissions (CSRC) passed Fair Disclosure regulation rules for publicly listed firms in 2006. In the same year, SZSE implemented fair information disclosure guidelines that require listed firms to disclose private meetings in annual reports. This requirement was updated in July 2012. Since then, all investor-relation events must be publicly disclosed in the required format within two trading days after the meeting, through a SZSE designated web portal.[36] Given this difference in sources, I collect data for 2006–2012 and 2012–2017 separately, and I cross-verify the overlapping period.

I obtain quarterly, semi-annual, and annual mandatory disclosure reports for all SZSE-listed firms during 2008–2012 from the exchange website. There are 21,514 files in total, each with a section for investor relation management activities. If private meetings occur during the disclosure period, information on the date, location, participants, and form are reported. I parse these entries from all reports to create a dataset, and I eliminate duplicate events if they are reported in more than one reports. For files composed in a format that does not allow the computer program to process, I manually collect information from them to ensure there are no missing records. A meeting is identified as *remote* if there are more than one city in the location field, if the meeting form is online or by phone, or if the meeting is organized by brokerage firms.

For each private meeting during 2012–2017, a typical report provides its date, location, and the names and employer institutions of all participating individuals, including both firm insiders and outsiders. In addition, the report also classifies the meeting into various categories, such as site visits, analyst-day meetings, online interactions, roadshows, and remote conference calls.[37] A summary of questions and answers during the meeting is also included in the report. From the designated Web portal, I extract reports that cover all events between 2012 and 2017. Beginning with 42,250 files, I carefully screen out 17,912 files that contain at least one mutual fund employee as a visitor. In a small fraction of files, reports are filed in non-text formats and cannot be processed by our program. For these files, I manually collect the relevant information. The categories and locations of meetings are also collected to differentiate site visits from other forms of communication. Specifically, a meeting is identified as a site visit if its category is either *research visit* or *site inspection* (or both), and no conference call is involved. With this requirement, 86,459 records are identified as site visits.[38]

Next, I combine these two datasets and exclude duplicate records in the overlapping period. I create

[36]See investor relation section (IRs): http://irm.cninfo.com.cn/szse/index_en.html (report filings are in Chinese language).

[37]For individual investors, the participant names are missing for many observations, so the actual number of individuals is understated.

[38]This criterion accurately identifies non-site meetings. For example, for 63,147 out of these 86,459 records, the location string explicitly mentions firm *office*, *conference room* or *reception room*. The remaining records typically include the typically contains location of the firm headquarters building.

a linktable between unique mutual fund families and various versions of their full names, abbreviations, nicknames, past names and the versions with different typos. Using this linktable, I identify visitors from report files and match them to their employer fund families. Individual visitor names are difficult to track, so to reduce noise, I drop all employee names, retaining only information about the firm, the fund family and the date of each record. If there are multiple visitors from the same fund family in a meeting, I keep only one record.

E Additional Discussions

Purpose of site visits. In general, it is unclear whether mutual fund managers visit firms to acquire private information for their portfolio management, or to monitor the firm management as shareholders. If the main purpose of the visits is related to corporate governance, then my empirical setting has a severe deviation from the theory to be tested.[39] Fortunately, a feature of my empirical setting largely alleviates this concern: Chinese mutual funds are rarely involved in corporate governance activities during the sample period.

In the 2016 China Securities Investment Fund Fact Book, it is documented that "...during year 2016, mutual funds voted 'For' on 12,185 corporate proposals, and only 101 'Against' (0.8%). Among all of their 12,338 votes, 97.3% votes are made through Internet."[40] This is in stark contrast to existing evidence from the US market.[41] A survey in 2014 also reveals that mutual funds are highly passive in corporate governance.[42] Specifically, it is "extremely rare" that mutual fund employees attend shareholder conferences, make inquiries or proposals, nominate corporate directors, or fight against controlling shareholders.

The main reason for Chinese mutual funds' limited involvement in corporate governance is likely their small ownership stake relative to controlling shareholders. Since the Chinese stock market is dominated by retail investors, the average (median) firm-level *aggregate* mutual fund ownership in my sample is only 5.9% (1.4%). Highly concentrated ownership structure of Chinese corporations limits the extent to which minority shareholders, including institutional investors, participate in corporate governance.[43]

Multi-office fund families. In my main analysis, I assume that each fund family has only one office location (as reported in the data), and all of its analysts and portfolio managers are based there. By interviewing practitioners from this industry, I learned that this is an innocuous assumption because most fund families do concentrate their core employees at one location. During more recent years, however, a few fund families allow a subset of portfolio managers to work at subsidiary offices that are located in different cities.

[39] Although information acquisition is a component of shareholder monitoring, the motivation and usage of information are different from assumptions in standard investment models.

[40] The document is available in Chinese language at http://www.amac.org.cn/researchstatistics/publication/zgzqtzjjynb/201801/t20180111_3356.html.

[41] For instance, see Matvos and Ostrovsky (2010) and Iliev and Lowry (2014).

[42] The survey is jointly performed by the Listed Companies Association of Shanghai, the Shanghai Asset Management Association, Shanghai Securities Association, and China Securities Investor Services Center. A summary of the survey results is available in Chinese language at http://news.stcn.com/2014/0214/11169569.shtml.

[43] Jiang and Kim (2015) report that, on average, the largest shareholder owns a third of the firm and the 5 largest own over half of the firm.

Although such cases should only weaken the power of my tests, to rule out this concern, I obtain a proprietary dataset from sell-side teams who directly serve the full universe of mutual fund managers in mainland China. This dataset contains information on whether a mutual fund family has subsidiary offices, and their locations, if any. When the subset of multi-office fund families are excluded from the sample, the estimated treatment effects are qualitatively similar with slightly larger magnitudes (see Table A.2).