DOI: 10.1111/fima.12354

ORIGINAL ARTICLE

Financial Management WILEY

Geographic proximity and price efficiency: Evidence from high-speed railway connections between firms and financial centers

Hao Gao[1] | Yuanyu Qu[2] | Tao Shen[3]

[1] PBC School of Finance, Tsinghua University, Beijing, China

[2] School of Banking and Finance, University of International Business and Economics, Beijing, China

[3] School of Economics and Management, Tsinghua University, Beijing, China

Correspondence
Yuanyu Qu, School of Banking and Finance, University of International Business and Economics, Beijing 100029, China.
Email: quyuanyu@uibe.edu.cn

Abstract

We study how geographic proximity to financial centers affects price efficiency. Using high-speed railway connections between firm cities and their nearest financial centers in China as exogenous shocks, we find stocks of connected firms are more efficiently priced than those of firms that are not connected. Consistent with our hypothesis, ease of travel has a stronger effect on firms that are closer to financial centers, smaller, have less institutional ownership and financial analyst coverage, and are not on the short sales list. Our paper highlights the importance of the geographic proximity of firms to financial centers on financial market efficiency.

KEYWORDS
Geographic proximity, High-speed railway, Information acquisitions, Price efficiency

JEL CLASSIFICATION
G10, G14

1 | INTRODUCTION

Location influences both information dissemination (Gurun & Butler, 2012) and information acquisition (Chen et al., 2021). Because agents make decisions based on their information, their geographic proximity to underlying security issuers plays a critical role in economics and finance. The literature has shown that location affects institutional investors' (Coval & Moskowitz, 2001) and individual investors' (Ivković & Weisbenner, 2005) investment decisions, financial analysts' forecasts (Malloy, 2005), bank lending (Degryse & Ongena, 2005), equity issuance (Loughran, 2008), mergers and acquisitions (Kang & Kim, 2008), and corporate investment (Giroud, 2013).

In this study, we provide evidence regarding the relation between a firm's geographic proximity to the nearest financial center and the informational efficiency of its stock prices. Price efficiency has significant implications for the capital market and real economy. More efficient prices better reflect the fundamentals of firms and help them make more informed investment and financing decisions (Chen et al., 2007). We hypothesize that the informational efficiency of stock prices is higher for firms that are closer to financial centers. Financial institutions are clustered in financial centers, so proximity to financial centers facilitates information dissemination and acquisitions. Accordingly, a short distance to financial centers attenuates information asymmetry between the firm and the financial markets and improves price efficiency.

We test our hypothesis using a sample of Chinese publicly listed firms from 2008 to 2016. China provides an ideal setting to investigate this issue. China covers a vast territory and the geographic distance constitutes a major cost in information acquisition and dissemination. In addition, the Chinese government imposes tight controls on the media and the Internet making some information difficult to acquire online. Moreover, China is experiencing rapid, yet unequal, development, and financial resources are concentrated in several metropolises. Due to these factors, the impact of location on the informational efficiency of stock prices is magnified.

We focus on the high-frequency measure of price efficiency that provides a more reliable estimate than long horizon measures in our context. The Chinese stock market is dominated by retail investors who trade excessively (Bailey et al., 2009) and the average annual turnover rate in China is typically more than twice that in the United States (Pan et al., 2016). Studies (Chordia et al., 2005) find that traders monitor the market closely and price discovery occurs primarily within one trading day. Due to the extreme turnover rate and the speed of price discovery, short-term measures are suitable to capture the relative informational efficiency of prices. Following previous studies (Boehmer & Wu, 2013; Cao et al., 2017), we assume that informationally efficient prices follow a random walk. We use intraday trade data from more than 1500 stocks from 2008 to 2016 to construct our measure as to how far transaction prices deviate from this benchmark. Using the method in Hasbrouck (1993), we separate the variation of a stock's efficient price from the variation of a pricing error by a vector autoregression model. The standard deviation of the pricing error measures the inefficiency of the stock price. We also consider an alternative proxy that is the absolute value of the quote midpoint return autocorrelation (Boehmer & Wu, 2013). A large absolute value of the autocorrelation indicates low price efficiency.

One difficulty in identifying the casual effect of geographic proximity on price efficiency is that firms with certain characteristics may be closer to financial centers and the observed relation can be driven by omitted variables. To establish the casual relation, we explore how shocks to geographic proximity affect price efficiency by studying the cities' connections to their nearest financial centers via the high-speed railway network in China. Construction of the high-speed railway network began in 2007. As of 2014, China had the world's longest and most active high-speed railway with connections among 81 cities and an annual ridership of 857 million. High-speed railway changes travel patterns profoundly, and a connected city usually experiences a significant increase in the number of train passengers (Lin, 2017). The shock to ease of travel affects different city pairs over time and generates cross-sectional variations.

We manually collect information about connections on the first high-speed railway between firm cities and their nearest financial centers and adopt a difference-in-difference research design. In particular, we compare the changes in the price efficiency of firms in cities that experience a shock to their transportation mode with the changes in price efficiency of firms in cities that do not experience a shock around the time when the high-speed railway connections are established. We expect the high-speed railway connection to attenuate the effect of distance and improve price efficiency.

The empirical tests support our hypothesis. We find that stock prices are more informationally efficient for firms that are closer to financial centers. Although geographic distance is time invariant, it generates important cross-sectional differences in the informational efficiency of stocks prices. More important, ease of travel attenuates the negative impact of geographic distance on price efficiency. Stock prices of firms in cities that are connected to the nearest financial center by the high-speed railway network are more informationally efficient than stock prices of firms that are not connected. This finding is not driven by the high-speed railway connection to other nearby important cities that are not financial centers. The results are also robust to the alternative proxy for price efficiency from Bartram & Grinblatt (2018).

The evidence from the high-speed railway connections helps us verify the channel through which location affects price efficiency. A firm's connection to a financial center facilitates information acquisition and dissemination. Because investors and financial analysts are more informed about firms connected to financial centers, their stock prices become more efficient. Next, we examine the effects conditional on the type of cities and further identify the mechanism behind our results.

We hypothesize that our findings are stronger for firms closer to the nearest financial center. For firms that are far from their nearest financial center, the benefits of the high-speed railway connection may be limited as air travel could be a superior transportation mode for long distance travel. To test this hypothesis, we adopt a triple difference type specification and find that after the connection, the price efficiency of proximate firms increases more than that of nonproximate firms. Relatedly, we also examine the differential effects of connections on the price efficiency of firms in large and small cities and find they have similar improvement in price efficiency.

We then examine the effects conditional on the characteristics of firms. Smaller firms tend to have less investor recognition, and their price discoveries may rely on soft information that is not included in financial reports. Because ease of travel could facilitate the acquisition of soft information, we hypothesize that our findings are stronger for these firms. Indeed, the results from the triple difference model indicate the effects of ease of travel on price efficiency are stronger for small firms. In addition, the shock to geographic distance changes the behavior of institutional investors and leads to more institutional ownership (Ellis et al., 2020). Because institutional ownership is associated with price efficiency (Boehmer & Kelley, 2009), we expect stocks with low institutional ownership to experience a significant improvement in price efficiency. We use mutual fund ownership as a proxy and find supporting evidence. Moreover, analyst coverage improves the information environment (Brennan & Subramanyam, 1995), and ease of travel facilitates information acquisition and production by financial analysts, especially for firms that are poorly covered. Consistently, we find that the effects of high-speed railway connections on the improvement of price efficiency are stronger for firms with lower analyst coverage. Finally, we exploit a condition in China that allows short sales for a designated list of firms. Short sales eligibility encourages information acquisition (Meng et al., 2020) and mitigates the negative impact of geographic distance on price efficiency. We expect the effects of high-speed railway connections on improving price efficiency should be weaker for firms on the short sales list and find supporting evidence. Taken together, these findings from the cross-sectional tests not only provide corroborating evidence for our main hypothesis, but also verify the mechanism behind the key results.

This paper contributes to the literature in several ways. First, it extends the literature on the determinants of price efficiency. Studies of the informational efficiency of stock prices largely focus on the role of institutional investors and short selling. Boehmer & Kelley (2009) find stocks with higher institutional ownership are priced more efficiently. Cao et al. (2017) find hedge funds tend to invest in stocks with low price efficiency and the price efficiency of those stocks is improved after the hedge fund holdings. Saffi & Sigurdsson (2010) use a global equity lending dataset and demonstrate that stocks with limited lending are associated with low price efficiency. Boehmer & Wu (2013) and Chang et al. (2014) find price efficiency is higher when short sellers are more active. We provide evidence that the geographic proximity of firms to financial centers is an important determinant of the informational efficiency of prices. More importantly, ease of travel attenuates the effect of distance and improves the price efficiency of firms with low institutional ownership and firms that cannot be shorted.

In addition, our paper contributes to the literature regarding the impact of location on financial markets. Studies suggest geographically proximate investors have an informational advantage. Teo (2009) finds evidence that geographically proximate hedge funds have a local informational advantage and outperform other hedge funds. Baik et al. (2010) suggest that local institutional ownership predicts future stock returns. Pirinsky & Wang (2006) present the impact of firms' headquarter locations on their returns. Hau (2001) uses a sample of firms in Germany to examine whether the local proximity of traders in financial centers to the corporate headquarters of the traded stocks affects traders' performance. The informational advantage related to location extends to other participants in the financial markets. Geographically proximate analysts issue more accurate forecasts and have a greater market impact than other analysts (Bae et al., 2008). Market makers closer to the firm's headquarters make more contributions to price discovery (Anand et al., 2011). Shive (2012) uses power outages as exogenous shocks to trading and examines the role of local investors on market efficiency. Our study complements the literature by demonstrating that the stock prices of firms that are connected to their nearest financial centers by high-speed railway networks are more informationally efficient. Thus, geographically proximate investors not only generate superior returns based on their informational advantage, but their informed trading also improves price efficiency.

Finally, our paper adds to the growing literature on the effect of transportation on financial markets. Herpfer et al. (2018) find that a reduction in travel time affects both new and existing borrowing relationships. The introduction of new airline routes affects venture capitalists' involvement with their portfolio companies (Mao et al., 2014; Bernstein et al., 2016), broadens firms' investor base, lowers their cost of equity (Da et al., 2021), affects mutual fund holdings (Ellis et al., 2020), and leads to an increase in plant-level investment from the headquarters and to total factor productivity (Giroud, 2013). Koudijs (2015, 2016) studies how private information is incorporated into prices and asset price volatility exploiting a natural experiment from the 18th century in which information flows were transmitted by sailing boats that were regularly interrupted by adverse weather conditions. Chen et al. (2021) find that mutual fund managers in China increase their site visits to a city after a direct high-speed railway connection is created. These studies all suggest ease of travel attenuates the effect of geographic distance on communication and information gathering. Our paper provides consistent evidence that access to a high-speed railway network enhances price efficiency resulting in an improvement in information acquisition and dissemination.

The rest of the paper is organized as follows. Section 2 describes the sample and descriptive statistics. Section 3 provides the main empirical analysis. Section 4 examines cross-sectional heterogeneity and the mechanism, and Section 5 provides our conclusions.

2 | DATA AND DESCRIPTIVE STATISTICS

We obtain accounting information, stock returns of publicly traded firms, mutual fund ownership, and financial analyst coverage data from the China Stock Market & Accounting Research (CSMAR) database. The sample period is from 2008 to 2016. We focus on the sample after 2008 for two reasons. First, the Chinese split-share reform was completed in 2007 and represents a major regime change in the development of the Chinese stock market (Liao et al., 2014). In addition, high-speed railways were introduced in 2008 and we rely on this shock to transportation to identify the impact of distance on price efficiency.

2.1 | Proximity to financial centers

Following the literature (Coval & Moskowitz, 1999, 2001; Loughran & Schultz, 2005; Anand et al., 2011), we define firm location as the city of the firm's headquarters. Major corporate decisions are made at their headquarters and they represent the information centers that connect firms with their suppliers, clients, and investors. We manually collect information about headquarter locations from the annual financial reports of companies.

To construct the proximity to financial centers, we consider three major metropolitan cities in China: Beijing in the north, Shanghai in the middle, and Shenzhen in the south. By the end of 2016, the gross domestic products (GDP) of these three cities in USD are \$377.26 billion, \$416.15 billion, and \$295.34 billion making them the second, the first, and the fourth largest cities in China, respectively. They not only play important roles in politics and economics, but also act as financial centers. Beijing is the capital city of China and contains all of the primary regulatory agencies including the People's Bank of China and the China Securities Regulatory Commission (CSRC). Shanghai and Shenzhen are home to the two major stock exchange markets in China: The Shanghai Stock Exchange (SHSE) and the Shenzhen Stock Exchange (SZSE).

Financial resources are lopsidedly distributed in these three cities and result in a conglomeration of institutional investors and sell-side financial analysts of brokerage firms.[1] Figure 1 provides the headquarter locations of the mutual fund families (in Panel A) and brokerage firms (in Panel B) in China at the end of 2016. A larger dot indicates a larger number of mutual fund families or brokerage firms in a city. Beijing, Shanghai, and Shenzhen are the headquarters for about 93% of the mutual fund families and about 57% of the brokerage firms. Therefore, most institutional investors and sell-side financial analysts are located in the financial centers. It supports our argument that being close to financial centers facilitates information dissemination and acquisitions.

To measure the physical proximity of firms to the three financial centers, we use a dummy variable that is equal to 1 if the firm is located within a certain distance to the nearest financial center and 0 otherwise. Alternative distances of 100, 200, 300, or 400 km are considered to construct this dummy variable. We exclude firms headquartered in these three cities to alleviate the concern that these firms may be fundamentally different and that our results are driven by local effects. We also exclude firms headquartered in Hainan Province, which is on a remote island, and firms in the financial industry.

2.2 | Measuring stock price efficiency

We focus on high-frequency measures of price efficiency using intraday transactions or quote data. High-frequency measures provide a more reliable estimate of price efficiency than the long horizon in our context. The Chinese stock market is dominated by retail investors that account for over 90% of the trading volume and over 99% of the trading accounts (Bailey et al., 2009). The literature (Mei et al., 2009; Xiong & Yu, 2011) indicates that these retail investors are vulnerable to behavioral bias and, as such, trade excessively (Pan et al., 2016). In addition, Chordia et al. (2005) confirm that most information is incorporated into prices within 30 min for stocks traded on the New York Exchange (NYSE). Thus, the extreme turnover rate in China combined with the speed of price discovery suggests that short-term measures are suitable to capture the relative informational efficiency of prices.

We obtain intraday price quote data from RESSET, a leading financial data vendor in China. RESSET provides detailed intraday security information for each trade including time, volume, price, and trade direction.[2] We eliminate trades with nonpositive price or volume and include only quotes that have positive depth. We also exclude quotes with nonpositive ask or bid prices or where the bid price is higher than the ask price. We require the difference between the bid and ask to be less than 25% of the quote midpoint and each security to have at least 60 trades per day and 200 transactions per year.

Our first measure of stock price efficiency is pricing errors variance that is first proposed by Hasbrouck (1993) and widely used in the literature (Boehmer & Kelley, 2009; Boehmer & Wu, 2013; Cao et al., 2017). To compute pricing error variance (PEV), assume that the observed log transaction price at time t, p_t, can be decomposed into an efficient price, m_t, and a transitory pricing error, s_t:

$$p_t = m_t + s_t, \tag{1}$$

where m_t is defined as the stock's expected value conditional on all available information at transaction time t and is assumed to follow a random walk. The pricing error s_t measures the deviation relative to the efficient price and is assumed to be a zero-mean covariance-stationary process. It captures noninformation-related market frictions and

[1] In China, investors trade by posting their own limit orders and effectively play the role of market makers to provide liquidity.

[2] Studies using NYSE's Trades and Quotes (TAQ) data typically rely on Lee & Ready's (1991) algorithm to assign trade directions.

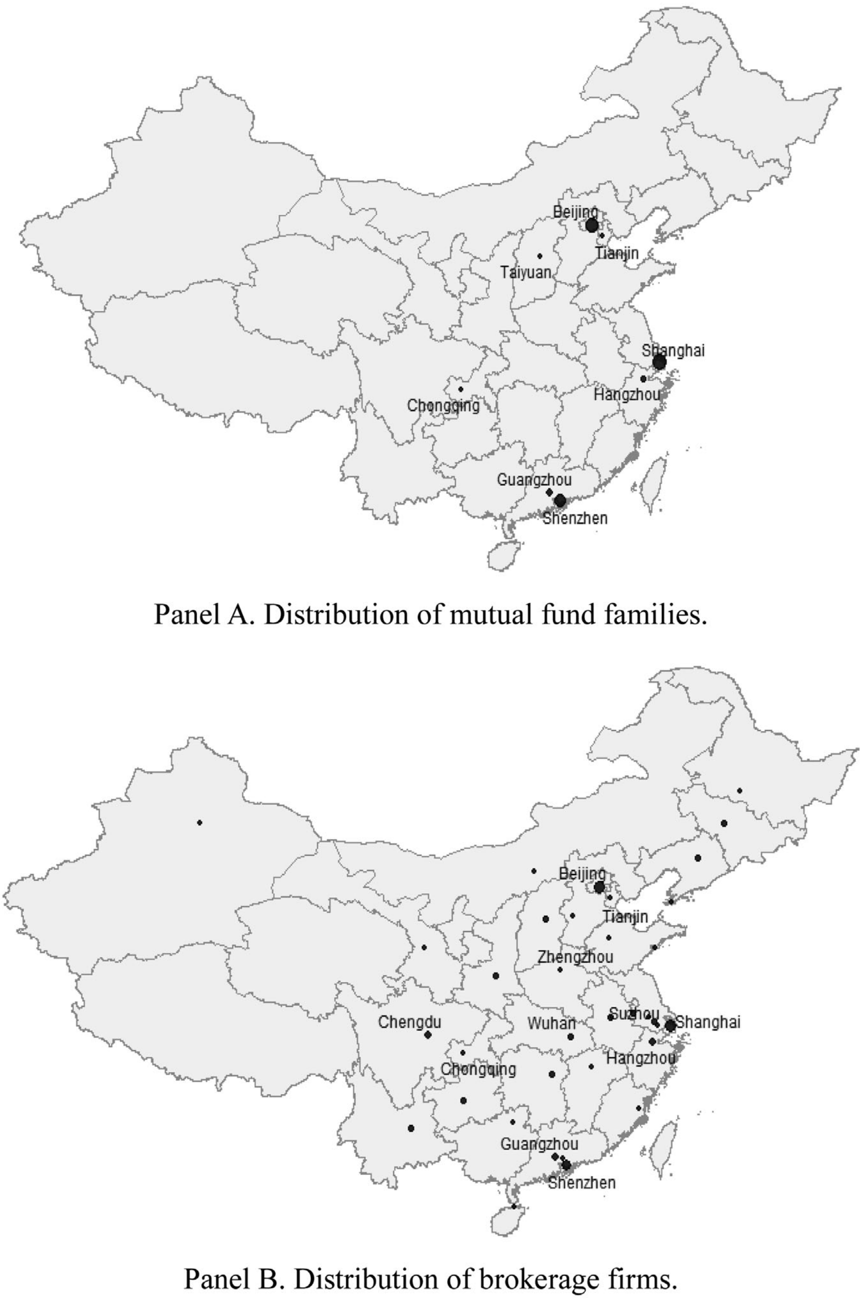

FIGURE 1 Distribution of financial institutions in China
Note: This figure depicts the headquarters locations of the mutual fund families (in Panel A) and brokerage firms (in Panel B) in China at the end of 2016. A larger dot indicates a higher number of mutual fund families or brokerage firms in a city. For figure brevity, the city of Sansha in Hainan Province is omitted

can be serially correlated or correlated with the innovation of efficient prices. Because the expected value of the deviations is zero, the standard deviation of the pricing error, σ_s, represents the magnitude of the deviations from the efficient price and is a measure of price efficiency. Instructions for estimating σ_s are presented in Appendix A. To make comparisons across stocks meaningful, σ_s is scaled by the standard deviation of price to control for cross-sectional differences in the return variance. Standardized PEV reflects the proportion of the deviation from the efficient price in the total variability of the observable transaction price process and is a measure of the informational efficiency of prices. A smaller PEV indicates a more efficient stock price.

Following the literature (Boehmer & Wu, 2013), our second measure of stock price efficiency is the absolute value of the quote midpoint return autocorrelation. The midpoint of the bid and ask price is the best estimate of the true value of the stock at any point in time. Therefore, if the prices are efficient, the midpoint returns should follow a random walk and their autocorrelations in either direction should be small. A large absolute value of the midpoint return autocorrelation indicates a low level of price efficiency. We calculate the return autocorrelation based on a nonoverlapping 20-min interval and use |AR20| to denote its absolute value. Results are similar if we choose a 15- or 30-min interval. For these two measures of price efficiency, we first calculate their daily value for each stock and then take the annual average.

2.3 | Other variables

To study the impact of proximity on price efficiency, we use control variables that are potentially associated with price efficiency: volume-weighted average price (VWAP), relative effective spread (RES), the absolute value of the order imbalance (OIB), mutual fund holdings (MFHD), net short selling flow (SHO), total firm assets (SIZE), return on assets (ROA), the leverage ratio (LEV), and Tobin's *Q* (*Q*).

We use VWAP to control for differences in price that can potentially affect trading costs and efficiency. Given the minimal bid–ask spread of one cent, a low stock price can be associated with a high percentage trading cost. A high trading cost deters investors and trading and the stock prices become stale resulting in a low level of price efficiency. We also use RES to directly measure trading costs. For each trade, RES is the absolute value of the difference between the execution price and the quote midpoint scaled by the quote midpoint, and we calculate the daily trade volume-weighted RES. High RES indicates a high trading cost and negatively affects price efficiency. We use OIB to control for trading pressure, which could affect price efficiency. Under excessive demand from either the buy side or the sell side, stock prices can deviate from their fundamental values and become less efficient. OIB is the absolute value of the difference between buyer-initiated trades and seller-initiated trades standardized by share volume. High OIB indicates high trading pressure and negatively affects the price efficiency. The literature finds short selling activities improve price efficiency (Saffi & Sigurdsson, 2010; Chang et al., 2014) and stocks with greater institutional ownership are priced more efficiently (Boehmer & Kelley, 2009; Hendershott et al., 2015). Accordingly, we include SHO and MFHD to control for the effects of short selling and institutional holdings. Finally, because some firm characteristics, such as firm size, profitability, leverage, and Tobin's *Q*, are potentially associated with price efficiency, they are included in the regressions. Variable definitions are provided in Appendix B.

Table 1 presents the descriptive statistics for the main variables. All continuous variables are winsorized at the top and bottom 1% level. We have about 13,000 firm-year observations. The number of firms in each year ranges from 986 in 2008 to 1959 in 2016, and they are from 241 cities (excluding the three financial centers). The average (median) distance between a firm and the nearest financial center is 522.01 km (407.90 km) with the 5th percentile of 84.76 km and the 95th percentile of 1223.47 km. PEV has an average of 0.09 and median of 0.08, whereas the average and median of |AR20| are 0.23 and 0.23, respectively. These numbers are quantitatively similar to those of the price efficiency measures based on U.S. data (Boehmer & Wu, 2013). Average mutual fund holdings represent 4.4% of firm outstanding shares. An important reason for the relatively low ownership is the high state ownership of publicly listed state-owned companies. These shares are transferable and counted as part of the public float, but are rarely available to the public and are seldom traded in the open market (Chen et al., 2021). The average and median of the net short selling are both effectively zero. This small number reflects the fact that China has only partially lifted its short selling ban (Jain et al., 2013; Chang et al., 2014) and only a number of stocks can be shorted.

TABLE 1 Descriptive statistics

Variable	Mean	SD	P5	P50	P95	N
PEV	0.09	0.05	0.05	0.08	0.18	13,028
|AR20|	0.23	0.02	0.20	0.23	0.27	13,026
DIS	522.01	469.49	84.76	407.90	1223.47	13,350
VWAP	16.58	13.68	4.59	12.32	44.01	13,028
RES	0.00	0.00	0.00	0.00	0.00	13,028
OIB	0.09	0.03	0.05	0.08	0.15	13,028
MFHD	0.04	0.07	0.00	0.02	0.18	13,355
SHO	0.00	0.00	−0.00	0.00	0.00	11,297
SIZE	7.98	1.17	6.31	7.87	10.18	13,355
ROA	0.06	0.06	−0.03	0.05	0.15	13,355
LEV	0.44	0.24	0.10	0.43	0.80	13,355
Q	2.78	2.15	1.05	2.09	6.86	13,114

Note: This table reports summary statistics for the main variables. The accounting and stock returns data are from the China Stock Market & Accounting Research (CSMAR) database. The intraday price quote data are from RESSET. Firm location information is manually collected. The sample period is from 2008 to 2016. DIS is the distance between a firm and its nearest financial center. Other variable definitions are in Appendix B. This table reports the mean, standard deviation, 5th percentile, 50th percentile, 95th percentile, and number of observations for the full sample.

In the untabulated analysis, we compare the price efficiency measures during the two subperiods of 2008–2012 and 2013–2016 and find that PEV and |AR20| are both smaller in the later sample period than in the earlier period. This holds for both the mean and median and suggests that stocks are more efficiently priced in 2013–2016. Because cities are connected to the high-speed railway network over time, ease of travel attenuates the effects of distance more strongly in the later sample period. Therefore, this pattern of price efficiency over time is consistent with our hypothesis. We also calculate the correlation coefficients for the main variables in the untabulated analysis. Consistent with our hypothesis, a high level of price efficiency is associated with a high trading price (low percentage trading cost), high institutional ownership, and active short selling activity, whereas large bid–ask spreads and one-sided trading pressure reduce price efficiency. It suggests that our measures are correct proxies for the underlying economic variables.

3 | GEOGRAPHIC PROXIMITY TO FINANCIAL CENTERS AND PRICE EFFICIENCY

3.1 | Univariate test

If geographic distance affects information acquisition and dissemination, price efficiency for firms close to financial centers should be higher than for firms far from financial centers. Before presenting the evidence from the multivariate analysis, we first conduct a univariate test to compare the price efficiency between proximate and nonproximate firms. A firm is considered a proximate (nonproximate) firm if the distance between the firm's city and the nearest financial center is within (more than) 300 km. Results are similar if alternative distances of 100, 200, or 400 km are used.

Table 2 presents the results. We find that both PEV and |AR20| are smaller for proximate firms than for nonproximate firms and the differences are statistically significant. This holds for both the average and the median. This result suggests that stocks of proximate firms are more efficiently priced. The difference in PEV and |AR20| in the univariate test also implies the impact of geographic proximity on price efficiency is prominent and an important feature in

TABLE 2 Proximate and nonproximate firms: Univariate test

	Proximate		Nonproximate		Difference	
Variable	Mean	Median	Mean	Median	Mean	Median
PEV	0.09	0.08	0.10	0.08	–0.00***	–0.00***
|AR20|	0.23	0.23	0.23	0.23	–0.00***	–0.00***
VWAP	17.63	13.23	15.84	11.71	1.79***	1.52***
RES	0.00	0.00	0.00	0.00	–0.00***	–0.00***
OIB	0.09	0.08	0.09	0.08	0.00	0.00
MFHD	0.04	0.02	0.05	0.02	–0.00**	–0.00
SHO	0.00	0.00	0.00	0.00	–0.00*	0.00
SIZE	7.89	7.80	8.04	7.92	–0.15***	–0.12***
ROA	0.06	0.06	0.05	0.05	0.01***	0.01***
LEV	0.41	0.39	0.47	0.46	–0.06***	–0.07***
Q	2.84	2.18	2.75	2.03	0.09**	0.15***

Note: This table reports the mean and median of the main variables for proximate and nonproximate firms and their difference between the mean and the median. A firm is defined as proximate (nonproximate) if the distance between the firm city and the nearest financial center is within (more than) 300 km. Variable definitions are in Appendix B. For the difference, ***, **, and * denote two-tailed statistical significance of the mean or median at the 1%, 5%, and 10% levels, respectively.

the data providing strong evidence for our hypothesis. Several firm characteristics are different between the proximate and nonproximate firms. This highlights the importance of controlling for these characteristics and dealing with omitted variables.

3.2 | High-speed railway connections and price efficiency

Although price efficiency is negatively associated with geographic distance as shown in Table 2, it is difficult to infer the casual impact in multivariate regressions. It is possible that firms with specific characteristics are close to financial centers and that omitted variables drive the results. To establish the casual relation, we study how shocks to geographic proximity affect price efficiency by exploiting the rapidly changing mode of transportation that significantly altered ease of travel between Chinse cities: the introduction of the high-speed railway. We consider the introduction of high-speed railway connections between firm cities and their nearest financial centers as shocks to ease of travel, and adopt a difference-in-difference approach to examine the causal impact of distance on price efficiency. Because ease of travel alleviates the geographic distance constraint by reducing the overall cost (time, effort, etc.), the high-speed railway connection facilitates information acquisition and dissemination. For example, Chen et al. (2021) find that mutual fund managers in China increase their site visits to a city after a direct high-speed railway connection is established. Thus, we expect high-speed railway connections to help attenuate the effect of distance and improve price efficiency.

The high-speed railway network in China underwent rapid development during our sample period. The State Council first released the Mid-to-Long Term Railway Development Plan in 2004 and issued a revised Plan in 2008. The Plan set the goal of a national high-speed railway grid composed of four north–south corridors and four east–west corridors with a budget of 4000 billion RMB (or about 571 billion USD). The purpose of this national project is to connect the major cities across provinces with faster transportation. The placement of high-speed railway routes, as well as the choice of cities to receive high-speed train stations, is based on factors including economic development, population, and resource distribution.

TABLE 3 High-speed railway connections to financial centers

Year	(1) Cities	(2) Firms
2008	0	0
2009	0	0
2010	12	247
2011	23	246
2012	21	118
2013	34	333
2014	22	98
2015	15	54
2016	9	70

Note: This table presents summary statistics for high-speed railway connections to the nearest financial centers. We manually collect information about the connections on the first high-speed railway between firm cities and their nearest financial centers. Column (1) reports the number of cities that are first connected to their nearest financial centers in each year. Column (2) provides the number of listed firms in these cities when the high-speed railway connections are first established.

Lin (2017) provides a comprehensive institutional background on the high-speed railway in China. According to a survey reported in the study, the average monthly income of high-speed train passengers is in the high-income group in China and about 40% of passengers travel for business purposes. Lin (2017) finds a clear increasing trend in ridership for high-speed railways. The railways mostly heavily used are medium-to-short lines connecting two cities, such as Beijing–Tianjin (113 km), Shanghai–Nanjing (271 km), and Guangzhou–Shenzhen (109 km). From 2010 to 2014, high-speed ridership increased from 300 to 830 million, and the percentage of ridership on different transportation modes increased from about 13% in 2010 to about 28% in 2014.

We manually collect information about connections on the first high-speed railway between firm cities and their nearest financial centers. Table 3 presents the summary statistics. Column (1) reports the number of cities that are connected to their nearest financial centers by high-speed train in each year. Column (2) reports the number of listed firms in these cities when the connections were first established.

In 2008 and 2009, there are no connected cities in our sample. In 2010, there are 12 newly connected cities and those cities have 247 publicly listed firms at that time. The number of connected cities varies over time and peaks at 34 in 2013. Variation in the number of firms over time follows that of the connected cities. We manually identify the existence and time of establishment of high-speed railway connections for the city pairs in our sample. The connection affects different firms over time and generates cross-sectional variations across cities. This provides an ideal setting to identify shocks to the ease of travel in attenuating information asymmetry and improving price efficiency.

We adopt the cohort-matching approach proposed by Gormley & Matsa (2011) to estimate the difference-in-difference model with multiple events. In particular, we compare the changes in the price efficiency of connected firms (treated group) and unconnected firms (control group) around the time when the high-speed railway connections are established. For each new connection, we construct a cohort of the treated and the control firms using firm-year observations for the 2 years before and the 2 years after the connection. Firms are required to be in the sample for the full 5 years (including the event year) around the event, and we obtain similar findings without this requirement. We then pool the data across cohorts and estimate the average treatment effect.[3] Specifically, we estimate the following

[3] This cohort matching approach also has the advantage of being easily extended into a triple difference type specification that is used in cross-sectional analyses in Section 4.

model:

$$Y_{i,c,t} = \beta_0 + \beta_1 Treat_{i,c} \times Post_{c,t} + \text{Controls}_{i,c,t-1} + \gamma_{i,c} + \omega_{c,t} + \varepsilon_{i,c,t}, \tag{2}$$

where i denotes the firm, c denotes the cohort, and t denotes the year. The dependent variable is one of two proxies for pricing efficiency. Treat is equal to 1 if the city of firm i in cohort c is connected to its nearest financial center by high-speed railway and 0 otherwise. Post is equal to 1 if the connection is established after year t in cohort c and 0 otherwise. The control variables include VWAP, RES, OIB, MFHD, SHO, SIZE, ROA, LEV, and Q. We include firm–cohort fixed effects, $\gamma_{i,c}$, to control for any fixed differences between firms, and we include year–cohort fixed effects, $\omega_{c,t}$, to control for any time trends. We allow the firm and year fixed effects to vary by cohort. If the high-speed railway connection facilitates the information acquisition and dissemination, we expect β_1, which captures the average treatment effect across multiple events, to be negative and significant.

Table 4 presents the results. In Column (1), we use PEV as the proxy for price efficiency. We find the interaction term has a negative and significant coefficient. Stock prices of firms in cities that are connected to the nearest financial centers by a high-speed railway network are more informationally efficient than stock prices of firms that are not connected. In particular, the connection results in a 9.7% increase in price efficiency, on average, after controlling for other factors. This suggests that geographic distance has a causal impact on price efficiency. It also helps us identify the channel through which location affects price efficiency as shown in Table 2. A connection to a financial center reduces travel costs and facilitates information acquisition and dissemination. Because investors are more informed about firms that are connected to the network, the stock price becomes more efficient.

The coefficients on the control variables generally have the expected signs. A higher RES indicates higher transaction costs deterring the trading of arbitrageurs and decreasing price efficiency. Stocks with more mutual fund holdings are priced more efficiently and the net shorting flow has a positive effect on price efficiency. These results confirm the previous finding that institutional investors and short sellers make prices informationally efficient.

In Column (1), we consider all firm-year observations of the control group that can be included when constructing a cohort. Although we control for a number of firm characteristics in Equation (2), it is still possible that the results are driven by the differences between the treated and the control firms. To address this concern, we match treated firms to control firms based on a number of firm characteristics including total assets, return on assets, leverage ratio, firm age, and state ownership. We then construct the cohorts using the matched sample. Column (2) reports the estimates from this matched sample with PEV as the proxy for price efficiency. Because of the matching, the sample size is reduced by 50%, but the main finding is largely unchanged. The coefficient on the interaction term is negative and significant and its magnitude is almost the same as in Column (1). Price efficiency is higher after high-speed railway connections were introduced between the firm city and the nearest financial center. In Columns (3) and (4), we use |AR20| as an alternative proxy for price efficiency and find similar and robust evidence.

We next show that the parallel-trend assumption underlying the difference-in-differences estimator is not violated. One concern with the difference-in-difference approach in Equation (2) is that the estimated treatment effect could be due to pretreatment differences in the characteristics of the treated and the control groups. To address this concern, we examine the dynamics of price efficiency around the new high-speed railway connections by adding two leads (before treatment) and two lags (after treatment) of the variable Post in Equation (2). The leads, Treat × Post (–2) and Treat × Post (–1), can control for pretreatment effects, whereas the lags, Treat × Post (+2) and Treat × Post (+1), can trace the treatment effects in the periods after the initial shock.

Table 5 presents the estimates based on this new specification. The dynamics of price efficiency around the high-speed railway connections largely support our hypothesis. We do not find a strong anticipatory effect. The interaction terms are generally not significant prior to the event, especially for the matched sample, so stock prices are not informationally more efficient before the connections. Alternatively, price efficiency is improved right after the connections and the significant effect of the lag interactions suggests the impact of the shock to the ease of travel is long lasting.

TABLE 4 Connections to financial centers and price efficiency

	(1)	(2)	(3)	(4)
	PEV		|AR20|	
	Full	Matched	Full	Matched
Treat × Post	−0.01***	−0.01***	−0.00***	−0.00**
	(−7.35)	(−5.86)	(−4.58)	(−2.54)
VWAP	0.00	0.00	0.00***	0.00***
	(0.95)	(0.97)	(7.45)	(7.64)
RES	14.76***	12.28***	4.74***	2.44***
	(8.42)	(7.48)	(6.36)	(5.08)
OIB	−0.01	−0.01	−0.03	−0.02**
	(−0.39)	(−0.51)	(−1.56)	(−2.15)
MFHD	−0.09***	−0.08***	−0.02**	−0.02
	(−6.23)	(−5.47)	(−2.39)	(−1.54)
SHO	−47.98***	−57.08***	6.07	3.38
	(−3.78)	(−4.43)	(0.87)	(0.69)
SIZE	−0.00	−0.00	−0.00***	−0.01***
	(−1.06)	(−1.30)	(−4.16)	(−8.97)
ROA	−0.07***	−0.08***	−0.01	−0.02**
	(−3.04)	(−4.42)	(−1.68)	(−2.65)
LEV	0.00	−0.00	0.00	0.01
	(0.09)	(−0.71)	(0.22)	(1.30)
Q	−0.00***	−0.00***	−0.00**	−0.00
	(−4.92)	(−3.56)	(−2.54)	(−1.19)
Year × cohort	Yes	Yes	Yes	Yes
Firm × cohort	Yes	Yes	Yes	Yes
N	18,799	9968	18,798	9967
Adj. R^2	0.60	0.60	0.21	0.23

Note: This table presents the results of the regression in Equation (2). The dependent variable is pricing error variance (PEV) in Columns (1) and (2) and |AR20| in Columns (3) and (4). |AR20| is the absolute value of the return autocorrelations calculated from the midpoints of the bid–ask spread quotes at nonoverlapping 20-min intervals (|AR20|). The accounting and stock return data are from the China Stock Market & Accounting Research (CSMAR) database. The intraday price quote data are from RESSET. The sample period is from 2008 to 2016. For each new connection, we construct a cohort of the treated and the control firms using firm-year observations for the 2 years before and the 2 years after the connection. In Columns (1) and (3), we use all firm-year observations of the control group that can be included when constructing a cohort. In Columns (2) and (4), we match treated firms to control firms based on a number of firm characteristics. Treat is equal to 1 if the city of firm in the cohort is connected to its nearest financial center by high-speed railway and 0 otherwise. Post is equal to 1 if the connection is established in a year in the cohort and 0 otherwise. Other variable definitions are in Appendix B. We include year × cohort and firm × cohort fixed effects. Standard errors are clustered at the industry level. *** and ** denote statistical significance at the 1% and 5% levels, respectively.

These results do not suggest that the parallel-trend assumption underlying the difference–indifferences estimator is violated.

Overall, the evidence in Tables 4 and 5 indicates that high-speed railway connections between firm cities and their nearest financial center have a significant impact on the informational efficiency of stock prices. The new mode of

TABLE 5 Connections to financial centers and price efficiency: Dynamic effects

	(1)	(2)	(3)	(4)
	PEV		|AR20|	
	Full	Matched	Full	Matched
Treat × Post (−2)	−0.00	0.00	−0.00	−0.00
	(−1.30)	(1.13)	(−1.54)	(−1.15)
Treat × Post (−1)	−0.01**	0.00	−0.00	0.00
	(−2.46)	(0.72)	(−0.54)	(0.61)
Treat × Post (0)	−0.01***	−0.00	−0.00***	−0.00***
	(−5.48)	(−1.39)	(−4.07)	(−3.37)
Treat × Post (+1)	−0.01***	−0.01**	−0.00	−0.00
	(−3.96)	(−2.44)	(−1.40)	(−1.63)
Treat × Post (+2)	−0.02***	−0.01***	−0.00**	−0.00**
	(−4.79)	(−7.04)	(−2.35)	(−2.63)
Controls	Yes	Yes	Yes	Yes
Year × cohort	Yes	Yes	Yes	Yes
Firm × cohort	Yes	Yes	Yes	Yes
N	18,799	9968	18,798	9967
Adj. R^2	0.60	0.60	0.21	0.23

Note: This table presents the results of the dynamic tests based on the model in Equation (2). The dependent variable is pricing error variance (PEV) in Columns (1) and (2) and |AR20| in Columns (3) and (4). |AR20| is the absolute value of the return autocorrelations calculated from the midpoints of the bid–ask spread quotes at nonoverlapping 20-min intervals (|AR20|). The accounting and stock return data are from the China Stock Market & Accounting Research (CSMAR) database. The intraday price quote data are from RESSET. The sample period is from 2008 to 2016. For each new connection, we construct a cohort of the treated and the control firms using firm-year observations for the 2 years before and the 2 years after the connection. In Columns (1) and (3), we use all firm-year observations of the control group that can be included when constructing a cohort. In Columns (2) and (4), we match the treated firms to the control firms based on a number of firm characteristics. Treat is equal to 1 if the city of the firm in the cohort is connected to its nearest financial center by a high-speed railway and 0 otherwise. Post (0) is equal to 1 if the connection is established in the event year in the cohort and 0 otherwise. We add two leads (before treatment) and two lags (after treatment) of the variable Post in Equation (2). Other variable definitions are in Appendix B. We include year × cohort and firm × cohort fixed effects. Standard errors are clustered at the industry level. *** and ** denote statistical significance at the 1% and 5% levels, respectively.

transportation alleviates the geographic distance constraint and facilitates information dissemination and acquisition leading to an improvement in the price efficiency.

4 | ADDITIONAL TESTS

In this section, we examine the robustness of our main findings and report the cross-sectional differences in the effect of high-speed railway connections on price efficiency to provide further evidence on the underlying mechanisms.

4.1 | Connections to other important cities

Figure 1 indicates financial resources are lopsidedly distributed in Beijing, Shanghai, and Shenzhen, and these three cities are the headquarters for more than 90% of mutual fund families and almost 60% of brokerage firms. It suggests the importance of the connection to financial centers in improving price efficiency. Nonetheless, the geographic proximity to other nearby important cities may also play a role.[4] In this subsection, we further examine whether the high-speed railway connection to other important cities affects the main findings.

We use two methods to define other important cities. First, we follow the plan of the central government on city development. In 2010, the Ministry of Housing and Urban-Rural Development of China issued the “National Urban System Plan” and designated five major cities, Beijing and Tianjin in the Bohai Economic Zone, Shanghai in the Yangtze River Delta Economic Zone, Guangzhou in the Pearl River Delta Economic Zone, and Chongqing in the West Triangle Economic Zone, as the National Central Cities. In 2016, Chengdu, Wuhan, and Zhengzhou were included in the National Central Cities. Therefore, the first definition provides six important cities in addition to the three financial centers. Additionally, we define the important cities based on their total annual GDP by the end of 2016. We consider the top 10 cities with the largest GDP. They include Shanghai, Beijing, Guangzhou, Shenzhen, Tianjin, Suzhou, Hangzhou, Chongqing, Chengdu, and Wuhan. The second method provides seven important cities excluding financial centers.

Next, we construct an indicator variable, *OtherConn*, that is equal to 1 after the firm’s city is connected to its nearest important city (excluding the financial centers) and 0 otherwise.[5] To avoid the multicollinearity problem, we exclude Tianjin and Guangzhou from the list of important cities.[6] We then add this indicator variable in Equation (2) and re-estimate the model.

Table 6 reports the results. In Columns (1), (3), (5), and (7), the important cities are Chongqing, Chengdu, Wuhan, and Zhengzhou. In Columns (2), (4), (6), and (8), the important cities are Suzhou, Hangzhou, Chongqing, Chengdu, and Wuhan. We find that the coefficient on the interaction term remains negative and significant. As such, the connection to other important cities does not drive our key findings. This finding highlights the importance of the connection to financial centers in improving the price efficiency. In contrast, the coefficient on *OtherConn* is insignificant. These results are robust to both proxies for price efficiency, both the full and matched samples of cohorts, and different definitions of important cities.

4.2 | Alternative proxy

In this subsection, we consider the mispricing measure from Bartram & Grinblatt (2018) as an alternative proxy for price efficiency and examine the robustness of our main findings. To evaluate stock market informational efficiency with minimal data snooping, Bartram & Grinblatt (2018) take an agnostic approach of fundamental analysis to estimate peer-implied fair values from all available accounting data. A higher level of mispricing, either underpriced or overpriced, represents a lower level of price efficiency. We follow Bartram & Grinblatt (2018) and construct the mispricing signal:

$$Mis_{i,t} = \frac{P_{i,t} - V_{i,t}}{V_{i,t}}, \tag{3}$$

where $P_{i,t}$ is the actual market equity value of firm *i* at date *t* and $V_{i,t}$ is the predicted intrinsic value based on accounting variables known by market participants at date *t*. We use 28 accounting variables to estimate the intrinsic value. The variable details are provided in Bartram & Grinblatt (2018, Appendix B). Mis measures the percentage difference between peer-implied fair value and actual market capitalization. Stocks with positive Mis are overvalued, whereas

[4] Note that we include the firm × cohort fixed effects in Equation (2). Thus, the physical proximity to the other nearest important city per se, which is invariant for a firm, is controlled by the fixed effects.

[5] For firms in these cities, *OtherConn* is equal to 1 after the firm’s city has the high-speed train stations and 0 otherwise.

[6] Tianjin is in the same economic zone with Beijing, and Guangzhou is in the same economic zone with Shenzhen. In the sample, a city is always connected to Beijing and Tianjin and to Shenzhen and Guangzhou at the same time. The inclusion of Tianjin and Guangzhou creates a multicollinearity problem for the treatment estimation.

TABLE 6 The impact of connections to other important cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	PEV				|AR20|			
	Full sample		Matched sample		Full sample		Matched sample	
	Central	GDP	Central	GDP	Central	GDP	Central	GDP
Treat × Post	−0.01***	−0.01***	−0.01***	−0.02***	−0.00**	−0.00**	−0.00**	−0.00***
	(−4.83)	(−5.24)	(−5.60)	(−6.19)	(−2.33)	(−2.13)	(−2.28)	(−3.06)
OtherConn	−0.00	−0.00	0.00	0.00	−0.00	−0.00	−0.00	−0.00
	(−1.27)	(−0.19)	(0.64)	(1.46)	(−0.59)	(−0.27)	(−1.26)	(−1.04)
VWAP	0.00	0.00	0.00	0.00	0.00***	0.00***	0.00***	0.00***
	(0.98)	(0.96)	(1.00)	(0.99)	(7.74)	(7.70)	(4.87)	(4.85)
RES	14.32***	14.38***	10.71***	10.70***	4.24***	4.26***	4.04***	4.06***
	(8.42)	(8.46)	(7.49)	(7.48)	(7.37)	(7.20)	(8.69)	(8.46)
OIB	0.00	0.00	0.04*	0.04*	−0.02	−0.02	−0.03**	−0.03**
	(0.10)	(0.12)	(1.87)	(1.85)	(−1.29)	(−1.28)	(−2.41)	(−2.40)
MFHD	−0.09***	−0.09***	−0.09***	−0.09***	−0.02**	−0.02**	−0.02	−0.02*
	(−6.90)	(−7.04)	(−5.47)	(−5.49)	(−2.42)	(−2.45)	(−1.72)	(−1.75)
SHO	−43.77***	−43.73***	−60.08***	−60.28***	8.24	8.26	5.66	5.86
	(−3.62)	(−3.57)	(−4.00)	(−4.00)	(1.29)	(1.30)	(1.12)	(1.20)
SIZE	−0.00	−0.00	−0.00	−0.00	−0.00***	−0.00***	−0.00***	−0.00***
	(−0.41)	(−0.39)	(−1.11)	(−1.09)	(−2.97)	(−3.01)	(−6.80)	(−7.01)
ROA	−0.08***	−0.08***	−0.08***	−0.08***	−0.01***	−0.01***	−0.01***	−0.02***
	(−3.85)	(−3.80)	(−5.45)	(−5.45)	(−3.03)	(−2.99)	(−3.40)	(−3.36)
LEV	−0.01	−0.01	0.00	0.00	0.00	0.00	0.01*	0.01*
	(−1.21)	(−1.15)	(0.15)	(0.14)	(0.41)	(0.42)	(1.87)	(1.92)
Q	−0.00***	−0.00***	−0.00***	−0.00***	−0.00**	−0.00**	−0.00**	−0.00**
	(−5.16)	(−5.19)	(−7.29)	(−7.14)	(−2.58)	(−2.62)	(−2.71)	(−2.85)
Year × cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm × cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	18,799	18,799	9968	9968	18,798	18,798	9967	9967
Adj. R^2	0.60	0.60	0.60	0.60	0.21	0.21	0.23	0.23

Note: This table presents the results of the regression in Equation (2) with the additional control variables. The dependent variable is pricing error variance (PEV) in Columns (1) to (4) and |AR20| in Columns (5) to (8). |AR20| is the absolute value of the return autocorrelations calculated from the midpoints of the bid–ask spread quotes at nonoverlapping 20-min intervals (|AR20|). The accounting and stock return data are from the China Stock Market & Accounting Research (CSMAR) database. The intraday price quote data are from RESSET. The sample period is from 2008 to 2016. For each new connection, we construct a cohort of the treated and the control firms using firm-year observations for the 2 years before and the 2 years after the connection. In Columns (1), (2), (5), and (6), we use all firm-year observations of the control group that can be included when constructing a cohort. In Columns (3), (4), (7), and (8), we match the treated firms to the control firms based on a number of firm characteristics. Treat is equal to 1 if the city of the firm in the cohort is connected to its nearest financial center by a high-speed railway and 0 otherwise. Post is equal to 1 if the connection is established in a year in the cohort and 0 otherwise. *OtherConn* is a dummy variable that is equal to 1 after a firm's city is connected to its nearest important city by a high-speed railway network and 0 otherwise. In Columns (1), (3), (5), and (7), the important cities are Chongqing, Chengdu, Wuhan, and Zhengzhou, which are defined by the central government's "National Urban System Plan." In Columns (2), (4), (6), and (8), the important cities are Suzhou, Hangzhou, Chongqing, Chengdu, and Wuhan and are defined according to a GDP ranking. Other variable definitions are in Appendix B. We include year × cohort and firm × cohort fixed effects. Standard errors are clustered at the industry level. *** and ** denote statistical significance at the 1% and 5% levels, respectively.

TABLE 7 Connections to financial centers and price efficiency: Alternative proxy

	(1) Full	(2) Matched
Treat × Post	−0.05**	−0.08***
	(−2.32)	(−3.78)
VWAP	0.00	0.00
	(0.76)	(0.06)
RES	5.40	−6.94
	(0.36)	(−0.56)
OIB	0.30	0.54
	(0.46)	(1.34)
MFHD	−0.16	−0.19
	(−0.66)	(−1.11)
SHO	−236.96	−237.67
	(−1.12)	(−1.02)
SIZE	−0.00	−0.03
	(−0.15)	(−0.94)
ROA	−0.82***	−0.60***
	(−7.26)	(−4.87)
LEV	−0.12	−0.00
	(−1.15)	(−0.02)
Q	−0.03**	−0.03***
	(−2.55)	(−4.10)
Year × cohort	Yes	Yes
Firm × cohort	Yes	Yes
N	18,719	9911
Adj. R^2	0.46	0.46

Note: This table presents the results of the regression in Equation (2). The dependent variable is the mispricing measure |Mis| from Bartram & Grinblatt (2018). The accounting and stock return data are from the China Stock Market & Accounting Research (CSMAR) database. The intraday price quote data are from RESSET. The sample period is from 2008 to 2016. For each new connection, we construct a cohort of the treated and the control firms using firm-year observations for the 2 years before and the 2 years after the connection. In Column (1), we use all firm-year observations of the control group that can be included when constructing a cohort. In Column (2), we match the treated firms to the control firms based on a number of firm characteristics. Treat is equal to 1 if the city of the firm in the cohort is connected to its nearest financial center by a high-speed railway and 0 otherwise. Post is equal to 1 if the connection is established in a year in the cohort and 0 otherwise. Other variable definitions are in Appendix B. We include year × cohort and firm × cohort fixed effects. Standard errors are clustered at the industry level. *** and ** denote statistical significance at the 1% and 5% levels, respectively.

stocks with negative Mis are undervalued. Because undervaluation or overvaluation represents mispricing, we take the absolute value of Mis as our measure of price efficiency (|Mis|).

We estimate Equation (2) with |Mis| as the dependent variable and report the results in Table 7. In both columns with different samples of cohorts, we find the coefficient of the interaction term is negative and significant. A high-speed railway connection to the nearest financial center significantly improves price efficiency. This finding is consistent with the results in Table 4 and provides further evidence for our main hypothesis.

4.3 | Different types of cities and price efficiency

In this subsection, we conduct tests to determine the city-level cross-sectional differences in the impact of high-speed railway connections on price efficiency and to provide further evidence regarding the underlying mechanisms. Loughran & Schultz (2005) find that urban firms whose headquarters are in the 10 largest metropolitan areas of the United States trade much more, have lower trading costs, are covered by more analysts, and are owned by more institutions than rural firms. This evidence indicates that it could be more difficult to obtain information on firms based in rural locations than on firms based in large metropolitan areas. In a similar spirit, we examine whether the impact of the high-speed railway connections on price efficiency is different across large and small cities and adopt a triple difference type specification as follows:

$$
\begin{aligned}
Y_{i,c,t} = {} & \beta_0 + \beta_1 \mathrm{Treat}_{i,c} \times \mathrm{Post}_{c,t} \times \mathrm{City}_j + \beta_2 \mathrm{Treat}_{i,c} \times \mathrm{Post}_{c,t} + \beta_3 \mathrm{City}_j \\
& + Controls_{i,c,t-1} + \mathrm{Industry} + \omega_{c,t} + \varepsilon_{i,c,t},
\end{aligned} \tag{4}
$$

where City_j denotes certain characteristics of the firm's city j, Industry denotes the industry fixed effects, and the other notations follow Equation (2).[7] We create a dummy variable, *Large_City*, that is equal to 1 if the population of a firm's city at the end of 2007 is above 5 million and 0 otherwise.[8] If the impact of the connection is more prominent for large cities, we expect that β_1 will be negative and significant.

Panel A of Table 8 presents the estimates. In all columns, the triple interaction term is negative, but insignificant for both proxies for price efficiency and both the full and matched samples of cohorts. Therefore, the stock price efficiency of firms in large cities is not improved significantly more than that of firms in small cities after high-speed railway connections to the nearest financial center.

The heterogeneity in the geographic proximity of firms to the nearest financial center may generate differential effects of the connections to price efficiency. As shown in Table 1, the average distance between a firm and the nearest financial center is 522.01 km, and it is about six times as long as that of the 5th percentile and less than half of that of the 95th percentile in the sample. For long distances, air travel could be a more convenient transportation mode than a train in terms of travel time. Thus, the impact of high-speed railway connections on price efficiency may be more prominent for firms closer to the nearest financial center.

To test this hypothesis, we estimate Equation (4) and the city characteristic is Proximity, which is a dummy variable that is equal to 1 if the distance between the firm city and its nearest financial center is within 300 km and 0 otherwise.[9] Panel B of Table 8 presents the estimates. In all columns, we find the triple interaction term is negative and significant. The stock price efficiency of firms that are close to their nearest financial center increases significantly more than that of firms that are far from financial centers after the high-speed railway connections. This evidence is consistent with idea that the high-speed train mainly facilitates short distance travel leading to a more prominent improvement in the price efficiency of proximate firms.

4.4 | Different characteristics of firms and price efficiency

Next, we examine the effects conditional on different characteristics of firms. We first explore how firm size affects our main finding. The literature (Grullon et al., 2004) indicates that larger firms tend to have greater investor recognition including a greater number of shareholders and more financial analyst coverage. Better investor recognition can

[7] We use industry fixed effects as firm–cohort fixed effects can subsume the cross-sectional differences among cities.

[8] The average city population at the end of 2007 is about 4.3 million people, and our results are similar if alternative cutoffs (e.g., 4 or 6 million) are used.

[9] Results are similar if we use 400 km to compute the indicator variable or the continuous variable of *–log(1+distance)*.

TABLE 8 Connections to financial centers and price efficiency: Conditional on the type of cities

Panel A. Large cities	**(1)**	**(2)**	**(3)**	**(4)**
	PEV		**|AR20|**	
	Full	**Matched**	**Full**	**Matched**
Treat × Post × *Large_City*	−0.00	−0.00	−0.00	−0.00
	(−0.62)	(−1.03)	(−1.49)	(−1.44)
Controls	Yes	Yes	Yes	Yes
Year × cohort	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
N	18,816	9974	18,815	9973
Adj. R^2	0.50	0.50	0.18	0.18
Panel B. Proximate cities	**(1)**	**(2)**	**(3)**	**(4)**
	PEV		**|AR20|**	
	Full	**Matched**	**Full**	**Matched**
Treat × Post × Proximity	−0.00*	−0.01***	−0.00***	−0.00***
	(−1.95)	(−3.11)	(−4.80)	(−5.37)
Controls	Yes	Yes	Yes	Yes
Year × cohort	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes
N	18,816	9974	18,815	9973
Adj. R^2	0.50	0.50	0.19	0.19

This table presents the results of the regression in Equation (4). The dependent variable is pricing error variance (PEV) in Columns (1) and (2) and |AR20| in Columns (3) and (4). |AR20| is the absolute value of the return autocorrelations calculated from the midpoints of the bid–ask spread quotes at nonoverlapping 20-min intervals (|AR20|). The accounting and stock return data are from the China Stock Market & Accounting Research (CSMAR) database. The intraday price quote data are from RESSET. The sample period is from 2008 to 2016. In Panel A, *Large_City* is a dummy variable that is equal to 1 if the population of a firm's city at the end of 2007 is above 5 million and 0 otherwise. In Panel B, Proximity is a dummy variable that is equal to 1 if the distance between the firm city and its nearest financial center is within 300 km and 0 otherwise. For each new connection, we construct a cohort of the treated and the control firms using firm-year observations for the 2 years before and the 2 years after the connection. In Columns (1) and (3), we use all firm-year observations of the control group that can be included when constructing a cohort. In Columns (2) and (4), we match the treated firms to the control firms based on a number of firm characteristics. Treat is equal to 1 if the city of the firm in the cohort is connected to its nearest financial center by a high-speed railway and 0 otherwise. Post is equal to 1 if the connection is established in a year in the cohort and 0 otherwise. Other variable definitions are in Appendix B. We include year × cohort and industry fixed effects. Standard errors are clustered at the industry level. *** and ** denote statistical significance at the 1% and 5% levels, respectively.

moderate the negative effect of geographic distance. In addition, soft information not included in the financial report could be important for small firms in their price discovery process, and ease of travel could facilitate the acquisition of soft information. Therefore, we expect that our finding should be stronger for small firms. To test this hypothesis, we construct an indicator variable, *Small_Firm*, that is equal to 1 if the firm size is below the sample median in the event year of the connection and 0 otherwise.[10] We use this indicator variable and estimate the following triple difference

[10] The indicator variable does not change within each cohort for a firm. Depending upon the sample distribution in the event years, a firm-year observation could be in a small firm group in one cohort and a large firm group in another cohort. The other indicator variables based on firm characteristics in the later analysis are defined similarly.

model:

$$Y_{i,c,t} = \beta_0 + \beta_1 \text{Treat}_{i,c} \times \text{Post}_{c,t} \times \text{Firm}_{i,c} + \beta_2 \text{Treat}_{i,c} \times \text{Post}_{c,t} + \text{Controls}_{i,c,t-1} + \gamma_{i,c} + \omega_{c,t} + \epsilon_{i,c,t}, \quad (5)$$

where $\text{Firm}_{i,c}$ denotes a certain firm characteristic. In this case, it is the *Small_Firm*.[11]

Panel A of Table 9 presents the results. In all columns, the estimate of the triple interaction term is negative and significant. It suggests the effects of ease of travel are mainly concentrated in small firms. After the firm city is connected to its nearest financial center through the high-speed train, the improvement in price efficiency for small firms is stronger than that of large firms. These results support our hypothesis and are consistent with the underlying mechanism.

Next, we investigate the role of institutional ownership that can lead to the differential effects of high-speed railway connections on price efficiency. Note that institutional ownership can be directly associated with price efficiency. Institutional investors prefer liquid stocks and these stocks have low transaction costs that facilitate arbitrage trading. Institutional investors may also have superior information about the firms and informed trading can increase price efficiency. Therefore, we include mutual fund ownership in Equation (2) as a proxy to control for the effects of institutional ownership. In addition to these direct effects, the shock to geographic distance can change the behavior of institutional investors in our setting. Ellis et al. (2020) find that introducing direct flights between a fund and a metropolitan statistical area leads to an increase in the fund's aggregate investment in firms in the area. Chen et al. (2021) confirm high-speed railway connections between mutual fund and firm cities increase site visits of fund managers to firms. Because ease of travel can lead to more institutional ownership and more information acquisition by institutional investors, we expect the effects of high-speed railway connections on price efficiency should be stronger for firms with lower institutional ownership.

We test this hypothesis by constructing an indicator variable, *Low_MFHD*, that is equal to 1 if mutual fund ownership is below the sample median in the event year of connection and 0 otherwise.[12] We estimate the triple difference model in Equation (5) with this indicator variable as the firm characteristics. Panel B of Table 9 presents the estimates. In Columns (1) and (2), we find the estimate of the triple difference is negative and significant for PEV in both the full and the matched sample. The price efficiency of firms with low mutual fund ownership is largely enhanced after the high-speed railway connection. This difference is consistent with our conjecture and provides corroborating evidence. In Columns (3) and (4), the dependent variable is |AR20|. Although the results are not as sharp as those in the first two columns, we still find a negative sign on the coefficient of the interaction term.

In a similar spirit, we also examine the role of analyst coverage. Both firm size and institutional ownership can be positively related to financial analyst coverage. Firms that are larger and those that have more institutional ownership attract more financial analysts. Research suggests that greater analyst coverage improves the information environment (Brennan & Subramanyam, 1995) and analysts' private communication with management is common and plays an important role in analyst research (Soltes, 2014; Cheng et al., 2016; Han et al., 2018). Ease of travel facilitates analyst visits to the firms and leads to greater information acquisition and production, especially for firms that are not well covered. Therefore, we expect the effect of high-speed railway connections on improving price efficiency should be stronger for firms with lower analyst coverage.

To test this hypothesis, we collect information about financial analyst coverage from the CSMAR database. Analyst coverage is defined as the total number of analyst research reports that cover a firm in a year and is adjusted by firm size and time. We construct an indicator variable, *Low_Coverage*, that is equal to 1 if analyst coverage is below the sample median in the event year of connection and 0 otherwise. We then estimate the triple difference model in Equation

[11] Because of the inclusion of firm–cohort fixed effects, the variable $\text{Firm}_{i,c}$ is omitted in Equation (5) as a control variable.

[12] Mutual fund ownership is highly associated with firm size. Following Hong, Lim, & Stein (2000) and Campbell, Hilscher, & Szilagyi (2008), we regress the mutual fund ownership of each firm on the firm size and time dummies, and use the residual as a proxy for mutual fund ownership that is independent from firm size. We use this proxy to compute *Low_MFHD*. We use the same approach to compute the size- and time-adjusted analyst coverage in the next test.

TABLE 9 Connections to financial centers and price efficiency: Conditional on the characteristics of firms

	(1)	(2)	(3)	(4)
	PEV		|AR20|	
Panel A. Firm size	**Full**	**Matched**	**Full**	**Matched**
Treat × Post × *Small_Firm*	−0.02***	−0.02***	−0.00**	−0.00***
	(−3.75)	(−4.00)	(−2.48)	(−3.29)
Controls	Yes	Yes	Yes	Yes
Year × cohort	Yes	Yes	Yes	Yes
Firm × cohort	Yes	Yes	Yes	Yes
N	18,799	9968	18,798	9967
Adj. R^2	0.60	0.60	0.21	0.23
	(1)	(2)	(3)	(4)
	PEV		|AR20|	
Panel B. Mutual fund ownership	**Full**	**Matched**	**Full**	**Matched**
Treat × Post × *Low_MFHD*	−0.01**	−0.01**	−0.00	−0.00
	(−2.29)	(−2.55)	(−0.83)	(−0.75)
Controls	Yes	Yes	Yes	Yes
Year × cohort	Yes	Yes	Yes	Yes
Firm × cohort	Yes	Yes	Yes	Yes
N	18,799	9968	18,798	9967
Adj. R^2	0.60	0.60	0.21	0.23
	(1)	(2)	(3)	(4)
	PEV		|AR20|	
Panel C. Analyst coverage	**Full**	**Matched**	**Full**	**Matched**
Treat × Post × *Low_Coverage*	−0.01**	−0.01***	−0.00*	−0.00*
	(−2.87)	(−2.96)	(−1.82)	(−1.77)
Controls	Yes	Yes	Yes	Yes
Year × cohort	Yes	Yes	Yes	Yes
Firm × cohort	Yes	Yes	Yes	Yes
N	18,799	9968	18,798	9967
Adj. R^2	0.60	0.60	0.21	0.23
	(1)	(2)	(3)	(4)
	PEV		|AR20|	
Panel D. Short sale	**Full**	**Matched**	**Full**	**Matched**
Treat × Post × *Off_Short*	−0.02***	−0.02***	−0.01***	−0.01***
	(−9.76)	(−10.06)	(−5.59)	(−6.38)
Controls	Yes	Yes	Yes	Yes
Year × cohort	Yes	Yes	Yes	Yes
Firm × cohort	Yes	Yes	Yes	Yes
N	18,799	9968	18,798	9967
Adj. R^2	0.60	0.60	0.21	0.23

(Continues)

TABLE 9 (Continued)

Note: This table presents the results from a series of triple difference type regressions in Equation (5). The dependent variable is pricing error variance (PEV) in Columns (1) and (2) and |AR20| in Columns (3) and (4). |AR20| is the absolute value of the return autocorrelations calculated from the midpoints of the bid–ask spread quotes at nonoverlapping 20-min intervals (|AR20|). The accounting and stock return data are from the China Stock Market & Accounting Research (CSMAR) database. The intraday price quote data are from RESSET. The sample period is from 2008 to 2016. In Panel A, *Small_Firm* is a dummy variable that is equal to 1 if the firm size is below the sample median in the event year of the connection and 0 otherwise. In Panel B, *Low_MFHD* is a dummy variable that is equal to 1 if the size- and time-adjusted mutual fund ownership is below the sample median in the event year of the connection and 0 otherwise. We regress the mutual fund ownership of each firm on the firm size and time dummies and take the residual to avoid grouping stocks by a characteristic that is highly correlated with size. We do the same for analyst coverage mentioned below. In Panel C, *Low_Coverage* is a dummy variable that is equal to 1 if the analyst coverage is below the sample median in the event year of the connection and 0 otherwise. In Panel D, *Off_Short* is a dummy variable that is equal to 1 if the stock is not short sales eligible in the event year of the connection and 0 otherwise. For each new connection, we construct a cohort of the treated and the control firms using firm-year observations for the 2 years before and the 2 years after the connection. In Columns (1) and (3), we use all firm-year observations of the control group that can be included when constructing a cohort. In Columns (2) and (4), we match the treated firms to the control firms based on a number of firm characteristics. Treat is equal to 1 if the city of the firm in the cohort is connected to its nearest financial center by a high-speed railway and 0 otherwise. Post is equal to 1 if the connection is established in a year in the cohort and 0 otherwise. Other variable definitions are in Appendix B. We include year × cohort and firm × cohort fixed effects. Standard errors are clustered at the industry level. *** and ** denote statistical significance at the 1% and 5% levels, respectively.

(5) with this indicator variable and report the results in Panel C of Table 9. In all columns, the triple interaction term is negative and significant. When firms are not well covered by financial analysts, the connection between the firm city and the nearest financial center largely improves price efficiency. This supports our conjecture that financial analysts are more likely to visit firms in connected cities resulting in greater information acquisition and production.

Finally, we examine the effects of high-speed railway connections on price efficiency for firms with different short selling status. The CSRC initiated a pilot program on March 31, 2010 by removing a short sales ban on a designated list of firms. The number of short sales eligible firms grew from 90 in the pilot program to about 900 by the end of 2016. The SHSE and SZSE established several requirements for firms to be included on the short sales list, and the list is adjusted each year. Beyond the direct effects that short selling activities have on improving price efficiency (Chang et al., 2014), Meng et al. (2020) find that negative media coverage of a firm increases after it is included on the short sales list. This suggests that short sales eligibility encourages information acquisition and mitigates the effect of geographic distance on price efficiency. Thus, we expect the effect of high-speed railway connections on the improvement in price efficiency should be weaker for firms on the short sales list.

We collect information about short sales status from the CSMAR database. We construct an indicator variable, *Off_Short*, that is equal to 1 if the stock is not short sales eligible in the event year of connection and 0 otherwise. We estimate the triple difference model in Equation (5) with this indicator variable and report the results in Panel D of Table 9. As expected, the interaction term is negative and significant across all four columns. The improvement in price efficiency after a high-speed railway connection is significantly higher for firms that cannot be shorted than firms on the short sales list. These findings support our hypothesis and are consistent with those in Panels A–C of Table 9. When a new mode of transportation alleviates the geographic distance constraint, the effects on price efficiency improvement are concentrated in firms that are more informationally opaque. This further corroborates the mechanism behind our main finding.

5 | CONCLUSION

The literature has long recognized the importance of geographic proximity in economics and finance. However, its role in determining the informational efficiency of stock prices has not received much attention. In this study, we examine how geographic proximity of firms to financial centers affects the efficiency of stock prices. Using the high-speed railway connections between firm cities and the nearest financial center in China as exogenous shocks, we find stock prices of connected firms are more efficiently priced than firms that are not connected. Our findings are consistent with the literature demonstrating that geographic proximity alleviates information asymmetry.

Our results are robust to alternative definitions of price efficiency. Consistent with our hypothesis, the effects of ease of travel on price efficiency are stronger for firms that are closer to their nearest financial center, smaller, have less institutional ownership and financial analyst coverage, and are not on the short sales list. Our study not only identifies geographic location as an important determinant of asset price efficiency, but also highlights the unique role of transportation and infrastructure construction in improving financial markets.

ACKNOWLEDGMENTS

The authors thank Bing Han (Editor), the anonymous reviewer, Kewei Hou, Qinghai Wang, Xiaoran Ni, and seminar participants at Tsinghua University and University of International Business and Economics for helpful comments and suggestions. Qu thanks the National Natural Science Foundation of China (grant ID 72003025) and University of International Business and Economics Outstanding Young Scholar Supporting Program (grant ID 20YQ13) for financial support. Shen thanks the Tsinghua University Initiative Scientific Research Program (grant ID 2019THZWJC17) and National Natural Science Foundation of China (grant ID 71603147) for financial support. All errors are of the authors.

REFERENCES

Anand, A., Gatchev, V. A., Madureira, L., Pirinsky, C. A., & Underwood, S. (2011). Geographic proximity and price discovery: Evidence from NASDAQ. *Journal of Financial Markets*, *14*, 193–226.

Bae, K. H., Stulz, R. M., & Tan, H. (2008). Do local analysts know more? A cross-country study of the performance of local analysts and foreign analysts. *Journal of Financial Economics*, *88*, 581–606.

Baik, B., Kang, J. K., & Kim, J. M. (2010). Local institutional investors, information asymmetries, and equity returns. *Journal of Financial Economics*, *97*, 81–106.

Bailey, W., Cai, J., Cheung, Y. L., & Wang, F. (2009). Stock returns, order imbalances, and commonality: Evidence on individual, institutional, and proprietary investors in China. *Journal of Banking and Finance*, *33*, 9–19.

Bartram, S. M., & Grinblatt, M. (2018). Agnostic fundamental analysis works. *Journal of Financial Economics*, *128*, 125–147.

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

Beveridge, S., & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle. *Journal of Monetary Economics*, *7*, 151–174.

Boehmer, E., & Kelley, E. K. (2009). Institutional investors and the informational efficiency of prices. *Review of Financial Studies*, *22*, 3563–3594.

Boehmer, E., & Wu, J. (2013). Short selling and the price discovery process. *Review of Financial Studies*, *26*, 287–322.

Brennan, M. J., & Subrahmanyam, A. (1995). Investment analysis and price formation in securities markets. *Journal of Financial Economics*, *38*, 361–381.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *Journal of Finance*, *63*, 2899–2939.

Cao, C., Liang, B., Lo, A. W., & Petrasek, L. (2017). Hedge fund holdings and stock market efficiency. *Review of Asset Pricing Studies*, *8*, 77–116.

Chang, E. C., Luo, Y., & Ren, J. (2014). Short-selling, margin-trading, and price efficiency: Evidence from the Chinese market. *Journal of Banking and Finance*, *48*, 411–424.

Chen, Q., Goldstein, I., & Jiang, W. (2007). Price informativeness and investment sensitivity to stock price. *Review of Financial Studies*, *20*, 619–650.

Chen, H., Qu, Y., Shen, T., Wang, Q., & Xu, D. (2021). The geography of information acquisition. *Journal of Financial and Quantitative Analysis*. https://jfqa.org/2021/03/15/the-geography-of-information-acquisition/

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

Chordia, T., Roll, R., & Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, *76*, 271–292.

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

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

Da, Z., Gurun, U. G., Li, B., & Warachka, M. (2021). Investment in a smaller world: The implications of air travel for investors and firms. *Management Science*, *67*(1), 417–435.

Degryse, H., & Ongena, S. (2005). Distance, lending relationships, and competition. *Journal of Finance*, *60*, 231–266.
Ellis, J. A., Madureira, L., & Underwood, S. (2020). The causal effects of proximity on investment: Evidence from flight introductions. *Journal of Financial and Quantitative Analysis*, *55*(6), 1978–2004.
Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *Quarterly Journal of Economics*, *128*, 861–915.
Gormley, T. A., & Matsa, D. A. (2011). Growing out of trouble? Corporate responses to liability risk. *Review of Financial Studies*, *24*, 2781–2821.
Grullon, G., Kanatas, G., & Weston, J. P. (2004). Advertising, breadth of ownership, and liquidity. *Review of Financial Studies*, *17*, 439–461.
Gurun, U. G., & Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. *Journal of Finance*, *7*, 561–598.
Han, B., Kong, D., & Liu, S. (2018). Do analysts gain an informational advantage by visiting listed companies? *Contemporary Accounting Research*, *35*, 1843–1867.
Hasbrouck, J. (1993). Assessing the quality of a security market: A new approach to transaction-cost measurement. *Review of Financial Studies*, *6*, 191–212.
Hau, H. (2001). Location matters: An examination of trading profits. *Journal of Finance*, *5*, 1959–1983.
Hendershott, T., Livdan, D., & Schürhoff, N. (2015). Are institutions informed about news? *Journal of Financial Economics*, *117*, 249–287.
Herpfer, C., Mjøs, A., & Schmidt, C. (2018). *The causal impact of distance on bank lending*. Working Paper.
Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, *55*, 265–295.
Ivković, Z., & Weisbenner, S. (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. *Journal of Finance*, *60*, 267–306.
Jain, A., Jain, P. K., McInish, T. H., & McKenzie, M. (2013). Worldwide reach of short selling regulations. *Journal of Financial Economics*, *109*, 177–197.
Kang, J. K., & Kim, J. M. (2008). The geography of block acquisitions. *Journal of Finance*, *63*, 2817–2858.
Koudijs, P. (2015). Those who know most: Insider trading in eighteenth-century Amsterdam. *Journal of Political Economics*, *123*(6), 1356–1409.
Koudijs, P. (2016). The boats that did not sail: Asset price volatility in a natural experiment. *Journal of Finance*, *71*(3), 1185–1226.
Lee, C. M., & Ready, M. J. (1991). Inferring trade direction from intraday data. *Journal of Finance*, *46*(2), 733–746.
Liao, L., Liu, B., & Wang, H. (2014). China's secondary privatization: Perspectives from the split-share structure reform. *Journal of Financial Economics*, *113*, 500–518.
Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from China's high speed railway system. *Journal of Urban Economics*, *98*, 98–123.
Loughran, T. (2008). The impact of firm location on equity issuance. *Financial Management*, *37*, 1–21.
Loughran, T., & Schultz, P. (2005). Liquidity: Urban versus rural firms. *Journal of Financial Economics*, *78*(2), 341–374.
Malloy, C. J. (2005). The geography of equity analysis. *Journal of Finance*, *60*, 719–755.
Mao, Y., Tian, X., & Yu, X. (2014). *Unleashing innovation*. Research paper, Kelley School of Business, Indiana University.
Mei, J., Scheinkman, J. A., & Xiong, W. (2009). Speculative trading and stock prices: Evidence from Chinese AB share premia. *Annals of Economics and Finance*, *10*, 225–255.
Meng, Q., Li, X., Chan, K. C., & Gao, S. (2020). Does short selling affect a firm's financial constraints? *Journal of Corporate Finance*, *60*, 101531.
Pan, L., Tang, Y., & Xu, J. (2016). Speculative trading and stock returns. *Review of Finance*, *20*, 1835–1865.
Pirinsky, C., & Wang, Q. (2006). Does corporate headquarters location matter for stock returns? *Journal of Finance*, *4*, 1991–2015.
Saffi, P. A., & Sigurdsson, K. (2010). Price efficiency and short selling. *Review of Financial Studies*, *24*, 821–852.
Shive, S. (2012). Local investors, price discovery, and market efficiency. *Journal of Finance*, *1*, 145–161.
Soltes, E. (2014). Private interaction between firm management and sell side analysts. *Journal of Accounting Research*, *52*, 245–272.
Teo, M. (2009). The geography of hedge funds. *Review of Financial Studies*, *22*, 3531–3561.
Xiong, W., & Yu, J. (2011). The Chinese warrants bubble. *American Economic Review*, *101*, 2723–2753.

How to cite this article: Gao H, Qu Y, Shen T. Geographic proximity and price efficiency: Evidence from high-speed railway connections between firms and financial centers. *Financial Management*. 2022;*51*:117–141. https://doi.org/10.1111/fima.12354

APPENDIX A: ESTIMATION OF PRICING ERROR VARIANCE

To estimate the pricing error, we estimate the following vector autoregression system with five lags:

$$\begin{aligned} r_t &= a_1 r_{t-1} + a_2 r_{t-2} + \cdots + b_1 \mathbf{x}_{t-1} + b_2 \mathbf{x}_{t-2} + \cdots + v_{1,t}, \\ x_t &= c_1 \mathbf{x}_{t-1} + c_2 \mathbf{x}_{t-2} + \cdots + d_1 r_{t-1} + d_2 r_{t-2} + \cdots + v_{2,t}, \end{aligned} \tag{A1}$$

where r_t is the difference in prices, p_t, and $\mathbf{x}_t$ is a vector of trade-related variables including a trade sign indicator, signed trading volume, and the signed square root of the trading volume to allow for concavity between prices and trades. $v_{1,t}$ and $v_{2,t}$ are zero mean, serially uncorrelated disturbances. The system Equation (A1) can be inverted to obtain its vector moving average representation that expresses the variables in terms of contemporaneous and lagged disturbances:

$$\begin{aligned} r_t &= a_0^* v_{1,t} + a_1^* v_{1,t-1} + a_2^* v_{1,t-2} .. + b_0^* v_{2,t} + b_1^* v_{2,t-1} + b_2^* v_{2,t-2} \cdots \\ x_t &= c_0^* v_{1,t} + c_1^* v_{1,t-1} + c_2^* v_{1,t-2} .. + d_0^* v_{2,t} + d_1^* v_{2,t-1} + d_2^* v_{2,t-2} \cdots \end{aligned} \tag{A2}$$

The pricing error under Beveridge and Nelson's (1981) identification restriction can be expressed as:

$$\begin{aligned} s_t &= \alpha_0 v_{1,t} + \alpha_1 v_{1,t-1} + \cdots + \beta_0 v_{2,t} + \beta_1 v_{2,t-1} + \cdots \\ \alpha_i &= -\sum_{k=i+1}^{\infty} a_k^*, \quad \beta_i = -\sum_{k=i+1}^{\infty} b_k^*. \end{aligned} \tag{A3}$$

Pricing error variance is then computed as:

$$\sigma_s = \sqrt{\sum_{j=1}^{\infty} [\alpha_j, \beta_j] Cov(v) \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix}}. \tag{A4}$$

APPENDIX B: VARIABLE DEFINITIONS

Name	Description	Definition
Proximity	Proximity to the nearest financial center	Geographic proximity of the firm to three financial centers (Beijing, Shanghai, and Shenzhen). It is a dummy variable that is equal to 1 if the firm is located within 100, 200, 300, or 400 km of the nearest financial center and 0 otherwise.
PEV	Pricing error variance	Based on Hasbrouck (1993). See Section 2.2.
|AR20|	Autocorrelation	Return autocorrelations calculated from the midpoints of the bid–ask spread quotes at nonoverlapping 20-min intervals.
VWAP	Volume-weighted average price	Daily volume-weighted average transaction price.
RES	Relative effective spreads	Daily volume-weighted relative effective bid–ask spread.
OIB	Absolute order imbalance	Absolute value of the daily share order imbalance standardized by share volume, where order imbalance is the difference between buyer-initiated trades and seller-initiated trades.
MFHD	Mutual fund holdings	Shares held by all mutual funds scaled by the number of shares outstanding at the end of the fiscal year.
SHO	Net short selling flow	Difference between the annual short flow and the annual short covering scaled by the annual trading volume. We multiply this measure by 100.

(Continues)

Name	Description	Definition
SIZE	Firm size	Logarithm of total assets.
ROA	Return on assets	Earnings before interest and taxes scaled by total assets.
LEV	Leverage	Total liability scaled by total assets.
Q	Tobin's *Q*	Market equity plus total book liability minus deferred taxes scaled by total assets.
|Mis|	Mispricing	Based on Bartram and Grinblatt (2018). See Section 4.2 for details.