Received: 1 August 2022 | Revised: 20 July 2024 | Accepted: 29 July 2024

DOI: 10.1111/jbfa.12830

JBFA

ARTICLE

Information acquisition and tax avoidance: Evidence from a natural experiment

Lyu Fan[1] | Caiyue Ouyang[2] | Jeffrey Pittman[3,4] | Jiacai Xiong[5] | Jun Yao[6]

[1]School of Economics and Management, China Agricultural University, Beijing, China

[2]School of Economics and Management, Beijing Jiaotong University, Beijing, China

[3]Faculty of Business Administration, Memorial University, St. John's, Newfoundland and Labrador, Canada

[4]Department of Accounting and Information Systems, Virginia Tech, Blacksburg, Virginia, USA

[5]School of Accounting, Jiangxi University of Finance and Economics, Nanchang, China

[6]Department of Accounting, Deakin University, Burwood, Victoria, Australia

Correspondence
Caiyue Ouyang, Beijing Jiaotong University, Beijing, China.
Email: caiyueouyang@bjtu.edu.cn.
Jiacai Xiong, Jiangxi University of Finance and Economics, Nanchang, China.
Email: xiongjiacai@jxufe.edu.cn

Abstract

Analyzing the launch of high-speed rail (HSR) services in China as a natural experiment, we identify a positive externality stemming from lower information acquisition costs: the reduction in firms' overinvestment in tax avoidance. Specifically, we find that outsiders undertake more corporate site visits and firms engage in less tax avoidance after the opening of HSR lines in the cities where these firms are located, leading to enhanced firm value. In another result consistent with expectations, we document that the impact of the introduction of HSR lines on tax avoidance is concentrated in firms in which insiders exhibit a high propensity to extract rents through aggressive tax strategies. Our results imply that more efficient transportation facilitates site visits and the acquisition of firm-specific information, particularly soft information. This improvement strengthens external monitoring, thereby limiting the ability of insiders to accumulate private benefits under the guise of tax avoidance that benefits all shareholders as the residual claimants.

KEYWORDS
agency problems, geographic proximity, high-speed rail, information acquisition, rent extraction, tax avoidance

JEL CLASSIFICATION
G34, H26, H54, O18

1 | INTRODUCTION

Grounded in an agency framework, recent theory and evidence suggest that although shareholders could benefit from the tax savings brought by firms' tax avoidance strategies, an overinvestment in tax avoidance undermines firm value given managers' rent-extraction incentives. These complex tax strategies enable insiders to conceal their opportunistic self-dealing such as earnings manipulation, perk consumption, related party transactions (*RPTS*) and other forms of resource diversion, which they can justify under the pretext that lowering corporate taxes benefits all shareholders as the residual claimants (e.g., Atwood & Lewellen, 2019; Bauer et al., 2020; Chen et al., 2010; Chen, El Ghoul, et al., 2022; Desai & Dharmapala, 2006, 2009a, 2009b; Desai et al., 2007; Khan et al., 2023). The rent-extraction-motivated tax avoidance largely stems from information asymmetry. However, there remains hardly any empirical evidence on whether reducing the costs associated with acquiring firm-specific information, especially soft information, curtails insiders' rent-extraction activities, thereby leading to a decrease in value-destroying corporate tax avoidance. In this study, we analyze this issue by exploiting a natural experiment involving the nationwide launch of high-speed rail (HSR) services in China as an exogenous shock to the cost of information acquisition.

Extensive prior work suggests that firm outsiders can elicit an information advantage from geographic proximity, which enables them to monitor firms more closely (e.g., Ayers et al., 2011; Baik et al., 2010; Ellis et al., 2020; Ivkovic & Weisbenner, 2005; Kubick et al., 2017; Malloy, 2005). Because developing transportation infrastructure shortens the space-time distance between firms and their stakeholders—and promotes their direct communication—it should also enhance external monitoring (e.g., Bernstein et al., 2016; Chemmanur et al., 2016). Face-to-face interactions are crucial for outsiders to learn about a company's strategic operations, assess the ethical standards of insiders, discern the true intentions behind various dealings and detect any covert, self-serving activities disguised as tax avoidance strategies (Chen, Ma, et al., 2022; Chen, Qu, et al., 2022; Cheng et al., 2016, 2019; Glaeser, 1999; Han et al., 2018; He et al., 2019; Ouyang et al., 2024; Storper & Venables, 2004). As such, we anticipate that the cost of resource diversion will rise with improvements to transportation systems, motivating insiders to refrain from relying on complex tax strategies to help hide their extraction of private benefits at the expense of outside shareholders.

HSR usually operates at speeds approaching 300 km/h, nearly twice as fast as traditional railway systems. Relative to other modes of transportation, HSR provides several advantages in terms of safety, cost, comfort, speed, capacity and punctuality (Givoni, 2006). China opened its first HSR line, between Beijing and Tianjin, in August 2008. Afterward, many more HSR lines have been constructed to connect the country's major cities. By December 31, 2018, the total length of HSR lines in China had reached 29,000 km, accounting for more than two-thirds of the world's total commercial HSR lines. The number of passengers had steeply risen from 7.3 million in 2008 to 2054.3 million in 2018 (China Statistical Yearbook 2019). Compared with the number of airline passengers, which was 610 million in 2018, HSR has indeed become the preferred transportation choice for inter-city travel in China. It follows that the large scale of the country's HSR system provides an opportune testing ground for examining its role in shaping the tax strategies that firms pursue. Moreover, given that decisions about when and where to construct HSR lines are made by the central government according to its transportation plan for the entire country, this event happens at different times in different cities, which is exogenous to firms' decisions and valuation (Cui & Li, 2019). We take advantage of this setting by relying on a staggered difference-in-differences (DID) design that helps dispel endogeneity threats to reliable inference to improve identification on whether insiders undertake complex tax strategies to facilitate the diversion of corporate resources.

Beyond the HSR context, analyzing data from China offers other advantages. Chinese firms often exhibit a significant divergence between control rights and cash flow rights. This disparity, accentuated by the country's relatively lax investor protection institutions (e.g., Jiang et al., 2010; La Porta et al., 1999, 2000), engenders a strong incentive for corporate insiders to divert firm resources in pursuing private benefits under the pretense of saving taxes (Lin et al., 2018). Consequently, this provides us with an opportune setting for examining firms' tax strategy decisions from an agency standpoint.

Desai and Dharmapala's (2006) complementarity theory held that outside investors suffer more information asymmetry when firms undertake more aggressive tax positions because insiders can exploit the complexity inherent in sophisticated tax planning to suppress their diversionary activities. However, injecting tension into our analysis, considerable prior evidence fails to support the complementarity between complex tax planning and insider diversion (e.g., Armstrong et al., 2015; Blaylock, 2016; Seidman & Stomberg, 2017).[1] Moreover, extensive research also implies that risk-averse managers may elect to underinvest in tax avoidance because pursuing aggressive tax positions not only costs managers time and effort but also leaves them susceptible to penalties and reputation loss (e.g., Chen & Chu, 2005; Gallemore et al., 2014). The enhanced communication that follows the development of transportation infrastructure may more closely align the interests of outside investors and managers, disciplining firms into implementing more value-enhancing tax planning (e.g., Khan et al., 2017; Lin et al., 2017).[2] Additionally, proximity may even afford firms an information advantage over tax authorities, which facilitates reducing their taxes (Kubick et al., 2017). Altogether, given the competing forces at work, the impact of the launch of HSR services on firms' tax positions distills to an empirical question.

In a DID analysis, our results lend support to the positive impact of reducing information acquisition costs on the value of tax strategies. We find that after the launch of HSR services, firms' tax avoidance subsides. Reflecting the first-order economic impact according to our coefficient estimates, firms' effective tax rates (ETRs) rise 1%, which accounts for 7% of the mean ETR in our sample, after the opening of HSR lines. We also find that the reduced tax avoidance after the introduction of HSR lines can be explained by the increase in the intensity of corporate site visits undertaken by institutional investors and financial analysts. This is consistent with the notion that site visits enable outsiders to better grasp firms' complex transactions, including their underlying motivations. Additionally, we report that the reduction in tax avoidance persists for at least 3 years afterward, which is robust to specifying alternative ETR proxies, conducting placebo tests, applying a stacked DID design with entropy balancing and employing an instrumental variable (IV) approach. In another result consistent with expectations, we document that the negative relation between the launch of HSR services and tax avoidance is concentrated in firms for which insiders exhibit a high propensity to exploit aggressive tax strategies to deprive outside investors of their fair share of corporate earnings. Importantly, in deepening the analysis to explore the underlying mechanism, we examine the mediation effect of income shifting on the relation between the launch of HSR services and tax avoidance. Income shifting between member firms within an affiliated group is a standard technique for orchestrating earnings management, tax avoidance and tunneling in China (Lo et al., 2010; Shevlin et al., 2012). We find that income shifting subsides after the opening of HSR lines, and the relation between the opening of HSR lines and tax avoidance is attenuated after controlling for income shifting. This evidence supports the narrative that enhanced external monitoring stemming from the launch of HSR services deters rent extraction through income-shifting activities, which, in turn, translates into firms taking less aggressive tax positions. Finally, we document that firm value rises with the fall in tax avoidance after the launch of HSR, corroborating the valuation effects of reducing information acquisition costs on tax strategies.

To help dispel the concern that competing explanations are responsible for the results, we conduct several additional tests. First, we examine whether the fall in tax avoidance stems from tax authorities imposing stricter enforcement. Next, we consider whether our core results spuriously arise from the government striving to cover the HSR construction costs with tax revenues from local firms. Additionally, we evaluate whether firms engage in less tax avoidance because financial constraints become more relaxed in the presence of HSR services, reducing firms' incentive to rely on aggressive tax strategies to conserve financial resources. Finally, we analyze whether firms pay higher ETRs simply because they have more investment opportunities in cities that charge them higher tax rates after the launch of HSR services. We find no empirical support for these alternative explanations.

[1] In the other direction, evidence empirically validating the complementarity theory includes Desai and Dharmapala (2006, 2009a), Desai et al. (2007), Kim et al. (2011), Chan et al. (2016) and Bauer et al. (2020).

[2] Indeed, prior work implies that stricter monitoring from institutional investors engenders higher corporate tax avoidance (Cheng et al., 2012; Khan et al., 2017; Chen et al., 2019).

We make several contributions to extant research. First, we enrich prior work examining firms' tax strategy decisions from an agency theory perspective. This literature suggests that insiders exploit complex tax planning to camouflage their diversion of corporate resources (e.g., Atwood & Lewellen, 2019; Bauer et al., 2020; Chen, El Ghoul, et al., 2022; Desai, 2005; Desai & Dharmapala, 2006; Khan et al., 2023). Although previous studies frequently link rent-extraction incentives with significant information asymmetry between insiders and outsiders, evidence on how information acquisition costs shape insiders' motivations to rely on tax strategies to facilitate their diversionary activities remains scarce. This is an interesting research question to explore given that even with reduced information acquisition costs, external stakeholders might still struggle to detect such self-dealing due to its inherent complexity. By focusing on an emerging market with severe rent-extraction issues and utilizing the HSR setting, we provide evidence implying that lowering the costs associated with information acquisition, particularly relating to soft information, leads to a steep reduction in detrimental tax avoidance practices. This finding not only sheds light on whether firms' tax strategies are sensitive to insiders' rent-extraction incentives but also suggests an efficient method for constraining aggressive tax avoidance driven by rent-extraction motives. As such, our results contribute to the public policy debates given that understanding these issues is pivotal for shaping regulatory measures and market practices that more effectively discipline corporate conduct.

Second, in response to calls for evidence on how firms' operating environments affect their tax planning (Dyreng & Maydew, 2018), we analyze the relation between geographic factors and firms' tax strategies. Geographic characteristics can play an integral role in shaping a firm's operating environment and, consequently, may influence their tax decisions. For example, Dyreng and Lindsey (2009) and Atwood and Lewellen (2019), Markle (2016) found that multinational companies locate their subsidiaries in tax havens to avoid taxes. More relevant to our purposes, Kubick et al. (2017) investigated how firms' geographic distance from tax authorities affects their tax avoidance. This research stream typically focuses on firms' tax strategies designed to generate tax savings and examines whether the trade-off between the risks and benefits of such strategies varies with geographic proximity. According to Kubick et al. (2017), proximity can even provide firms with an informational advantage that facilitates tax avoidance. Diverging from prior studies, we examine the impact of geographic proximity on firms' incentives to extract rents under the guise of tax avoidance that benefits all shareholders. In documenting that proximity enhances external monitoring and curtails rent-extraction-motivated tax avoidance, we contribute to extant research by deepening our understanding of the interactions between geographic factors and tax planning decisions.

Third, we advance emerging research on micro-level outcomes arising from transportation infrastructure development. Research on the consequences of such development primarily focuses on its macroeconomic implications (e.g., Donaldson & Hornbeck, 2016). More recently, studies have begun to explore how improving transportation infrastructure can affect firms' activities at the micro-level, such as their inventory management (Cui & Li, 2019), production networks (Bernard et al., 2019), investment and productivity (Giroud, 2013), innovation (Bernstein et al., 2016; Hou et al., 2018), audit quality (He et al., 2019) and trade credit (Ouyang et al., 2024). Our analysis broadens this research stream by examining the importance of transportation infrastructure development to corporate decisions from an agency theory perspective. Constructive from an identification standpoint, large-scale investment in transportation systems usually occurs in developing countries where information asymmetry is severe and corporate governance is poor. Besides its role in accelerating economic growth, our research implies that transportation infrastructure development engenders a positive externality in the form of alleviating agency costs evident in firms' tax avoidance.

Finally, our evidence helps inform the public policy discourse on reforms to country-level tax institutions. Naturally, corporate tax revenue is a major source of government income. Indeed, Dyreng et al. (2008) documented that aggregate cash payments over a 10-year period amount to one third of aggregate pre-tax book income over this timeframe. However, many countries, particularly in emerging markets, are beset by poor tax institutions that facilitate serious tax evasion (e.g., Cai & Liu, 2009; Lin et al., 2018). For example, firms in China paid 169 billion RMB (equivalent to US$24 billion) in overdue taxes in 2016 alone (China Tax Audits Yearbook 2017). Our results imply that developing its transportation infrastructure enables the government to constrain tax revenue losses. Although it would be premature at

this early stage to reach firm policy prescriptions, our evidence lends some preliminary support that the government's tax revenues will rise when they make such investments.

The rest of this paper is organized as follows. Section 2 motivates the institutional setting and relies on prior theory and evidence in developing the hypotheses. Section 3 describes the sample and outlines the research design. Section 4 covers the empirical results, and Section 5 concludes.

2 | INSTITUTIONAL SETTING AND HYPOTHESES DEVELOPMENT

2.1 | Tax avoidance and rent extraction in China

China's central government instituted a tax-sharing system in 1994. Under this system, corporate taxes are classified into central taxes and local taxes that are separately collected by the central government and local governments, respectively. For long-term tax revenues, many local governments offer outside companies explicit or implicit tax benefits to attract investment (Wu et al., 2007). The central government also relies on tax policies to persuade investors to invest in specific regions and industries. As such, there is wide variation in statutory tax rates, exemptions, tax credits (e.g., recycling, environmental protection) and reductions (e.g., to high-tech firms and new firms operating in impoverished areas) (Cai & Liu, 2009). Such variation affords firms ample tax avoidance opportunities (Lin et al., 2018).

Typical tax avoidance methods include misreporting sales revenue and expenses, abusing tax credits, manipulating earnings and transferring profits (Cai & Liu, 2009; Lin et al., 2018). Among these methods, transferring profits between group members within a pyramid structure is the most attractive because it does not affect the group's consolidated pre-tax income, which is usually difficult for tax authorities to detect and challenge such transactions (Lin et al., 2018). In China, discipline against insiders diverting corporate resources for their own benefit is minimal given the weak governance, lax investor protection and concentrated control rights (e.g., Jiang et al., 2010; La Porta et al., 1999, 2000). Aggressive tax avoidance strategies not only help insiders generate more cash flows for diversion purposes but also enable such misbehavior to fly under the radar by undermining financial reporting transparency (Bauer et al., 2020). In short, aggressive tax avoidance and rent extraction are both prevalent and closely associated with each other in China. Consequently, our setting suits analyzing whether a shift in rent-extraction incentives affects tax avoidance.

2.2 | Background on HSR in China

In the past three decades, China has enjoyed rapid economic growth. However, the development of its transportation infrastructure has failed to keep pace, impeding further economic development. Until recently, the speed of commercial train service was limited to a moderate level because most trains had to share tracks, which had already been overloaded, with freight trains (Lawrence et al., 2019). In response, the Chinese government decided to build HSR to connect provincial capitals and other large cities. In 2004, the State Council announced its plans to construct a "Four-Horizontal and Four-Vertical" (four east–west and four north–south lines) HSR corridor by 2020. This plan was extended to the "Eight-Horizontal and Eight-Vertical" corridor in 2016. Compared with most other countries, China is more suitable for HSR development. For starters, it covers a large geographic area (9.6 million km^2) with long distances separating major cities. Additionally, China has substantial population density (141 people/km^2). Finally, the construction of HSR is backed by a strong economy (GDP per capita = US$7590 in 2017) (Lawrence et al., 2019). Accordingly, it is not surprising that China has gradually established the largest HSR network in the world.

Since China launched its first HSR, the Beijing–Tianjin line, in 2008, its HSR network has quickly expanded. By 2018, its network had covered 29 out of 32 provinces, and the length of HSR lines had reached 29,000 km. It is expected that the total length of HSR lines will reach 38,000 km by 2025, linking all large cities. At that stage, the HSR network will

cover over 90% of the population, and travel time across China will be reduced by about 70% (Cui & Li, 2019). The rapid development of HSR improves the mobility of capital, labor, technology and information.

2.3 | Hypothesis development

Extensive prior work on the impact of geographic proximity has shown that analysts, auditors and investors usually have more information about firms located nearby, reflecting lower information acquisition costs (Baik et al., 2010; He et al., 2019; Ivkovic & Weisbenner, 2005; Malloy, 2005). The information advantage stemming from geographic proximity enables regulators, auditors, directors and institutional investors to more closely monitor firms (Ayers et al., 2011; Baik et al., 2010; He et al., 2019; Knyazeva et al., 2013; Kubick et al., 2017). Moreover, firms located near investors or financial centers attract cheaper equity financing costs (Degryse & Ongena, 2005; El Ghoul et al., 2013). Collectively, extant research implies that geographic proximity significantly influences information acquisition costs and the degree of information asymmetry that investors experience.

Given that the launch of HSR services drastically shortens the space-time distance between firms and their stakeholders, we expect it to reduce information acquisition costs and narrow information asymmetry. For starters, it improves the flow of information and personnel between cities (Dong et al., 2020; Lin, 2017), thereby reducing the time and costs for outsiders to acquire firm-specific information. Additionally, it facilitates site visits and face-to-face communication by analysts, auditors and investors (Chen, Ma, et al., 2022; Chen, Qu, et al., 2022; Cheng et al., 2016, 2019; Glaeser, 1999; Han et al., 2018; He et al., 2019; Liu et al., 2017; Ouyang et al., 2024; Storper & Venables, 2004). These interactions enable outsiders to better understand the true motivations for firms' transactions, translating into stricter external monitoring. Finally, reflecting the lower monitoring costs, the ensuing shorter travel time under HSR also helps attract more analysts, independent directors and institutional investors to the firm, further enhancing information production (Loughran & Schultz, 2005; O'Brien & Tan, 2015).

Some prior research implies that insiders orchestrate aggressive tax strategies for rent-extraction purposes. In dissecting several high-profile corporate governance failures such as Enron and Xerox, Desai (2005) provided anecdotal evidence that managers exploit complex tax planning to siphon firm resources. Desai and Dharmapala (2006) found that incentive compensation reduces the level of tax sheltering by alleviating conflicts of interest between shareholders and managers. Moreover, Desai et al. (2007) and Desai and Dharmapala (2009a) found that aggressive tax planning facilitates insiders' rent extraction when lax corporate governance engenders organizational complexity and financial opacity. Reinforcing this evidence, Frank et al. (2009) reported that firms deploying aggressive tax strategies tend to concurrently manage their book earnings upward. Atwood and Lewellen (2019) reported a positive relation between tax avoidance and management diversion for firms based in countries with weak investor protection but incorporated in tax havens. Additionally, Chen, El Ghoul et al. (2022) and Khan et al. (2023) showed that enhanced corporate governance reduces firms' tax avoidance, whereas weaker governance structures lead to more aggressive tax planning. These findings are based on analyses using cross-listings in the United States and director and officer liability as proxies for corporate governance, respectively.

Kim et al. (2011) identified another negative implication of tax avoidance: Because managers are eager to hide bad news through complex tax strategies, firms undertaking more aggressive tax positions exhibit higher stock price crash risk. By showing the actual mechanisms through which insiders extract private benefits under the guise of tax avoidance, recent research provides direct evidence for the complementarity theory. Chung et al. (2019) found that insiders realize higher purchase profitability from trading in firms with high levels of tax avoidance arising from the information advantage they elicit from poor accounting transparency. Using path analysis, Bauer et al. (2020) documented that the additional cash flows and increased opacity due to aggressive tax planning provide controlling shareholders in China with both the incentive and opportunity to orchestrate tunneling activities.

If lower information acquisition costs and an effective monitoring environment constrain insiders' self-dealing, it follows that their incentive to engage in aggressive tax strategies will subside because the costs for insiders to extract

private benefits using tax avoidance will rise in this situation. Lending support to this intuition, Lanis and Richardson (2011) found that tax avoidance is lower in firms with more outside members on the board. In a similar vein, Khurana and Moser (2013) reported that firms held by long-term institutional investors engage in less tax avoidance, which they attribute to concerns about opportunistic managerial behavior and the opacity stemming from tax avoidance activities. Additionally, more recent research implies that firms undertake more tax avoidance after an exogenous fall in analyst coverage (e.g., Allen et al., 2016; Chen & Lin, 2017; Chen et al., 2018).

Accordingly, we expect that the decrease in information acquisition costs and enhanced monitoring that ensues after the launch of HSR services will motivate insiders to reduce their rent-extraction activities, resulting in their firms exhibiting lower tax avoidance. We state our first prediction as follows:

H1. *Ceteris paribus*, the level of firms' tax avoidance declines after the launch of HSR services.

We expect that the impact will be stronger for firms in which insiders have wider scope to extract private benefits before the launch of HSR services because these firms are apt to have a high level of rent-extraction-induced tax avoidance in the first place. Consequently, we formulate our second hypothesis as follows:

H2. *The negative association between the launch of HSR services and the level of firm's tax avoidance is stronger for firms in which insiders have a high rent-extraction propensity*.

Our empirical predictions are rooted in the premise that firm insiders engage in aggressive tax strategies partly for rent-extraction purposes. However, recent research calls this premise into question. In large-sample analysis, Blaylock (2016) failed to find a perceptible link between tax avoidance and opportunistic managerial behavior. Seidman and Stomberg (2017) provided an alternative explanation for the findings in Desai and Dharmapala (2006), casting doubt on the validity of the relation between tax avoidance and rent extraction. Additionally, there is another stream of research focusing on the incentives of insiders to underinvest in tax avoidance. Although undertaking more aggressive tax positions benefits all shareholders as the residual claimants, the reality is that this not only costs managers in time and effort but also leaves them vulnerable to penalties and reputation loss (e.g., Badertscher et al., 2013; Chen & Chu, 2005; Gallemore et al., 2014). Accordingly, managers may elect to engage in overly conservative tax strategies unless they receive proper compensation (Armstrong et al., 2012; Phillips, 2003). Stricter monitoring and more efficient communication in the wake of HSR services may more closely align the interests of outside investors and managers, disciplining firms into pursuing more aggressive tax planning. Consistent with this view, prior work has shown that closer monitoring by institutional investors leads to an increase in firms' tax avoidance (Chen et al., 2019; Khan et al., 2017). Moreover, improving transportation infrastructure may even enable firms to enjoy an information advantage over tax authorities, facilitating tax avoidance (Kubick et al., 2017). Altogether, how the launch of HSR services affects firms' tax planning amounts to an empirical issue.

3 | SAMPLE, DATA AND RESEARCH DESIGN

3.1 | Sample

We collect financial data from the China Stock Market and Accounting Research database. Data on corporate site visits come from the Chinese Research Data Services platform. We manually collect information about the opening of HSR lines from the website of China's National Railway Administration (www.12306.cn). The sample period begins in 2005, which is 3 years before 2008 when the first HSR line was opened.[3] Our initial sample includes all firms listed

[3] In a robustness check, we verify that our core results are nearly identical when we set 2000 and 2003 as alternative starting years.

on the Shanghai Stock Exchange or the Shenzhen Stock Exchange, amounting to 52,500 firm-year observations from 2005 to 2022. Next, we exclude as follows: (i) 364 firm-years from the financial industry; (ii) 5970 observations with ETRs larger than 1 or less than 0[4]; (iii) 2889 observations with negative pre-tax income; (iv) 20,548 firm-years from the 4 municipalities (i.e., Beijing, Tianjin, Shanghai and Chongqing) and 28 provincial capital cities to confront any lingering endogeneity concerns[5]; and (v) 4895 firm-years with missing values for the control variables used in our analyses.[6] Our final sample contains 17,834 firm-year observations representing 2,210 unique firms.

3.2 | Research design

To test our hypotheses, we develop a series of measures for the intensity of corporate site visits, tax avoidance and insiders' rent-extraction propensity. In Appendix A, we provide detailed variable specifications.

3.2.1 | Measures for the intensity of corporate site visits

Corporate site visits. Consistent with prior research (Cao et al., 2022; Cheng et al., 2016, 2019; Han et al., 2018; Jiang & Yuan, 2018), we measure the intensity of corporate site visits with the frequency of visits by institutional investors (*Visit_Inst*), financial analysts (*Visit_Anlyst*), individual investors (*Visit_Indiv*) and news media (*Visit_media*).

3.2.2 | Measures for tax avoidance

Among the proxies for tax avoidance that are widely applied in prior work, we primarily focus on the three that suit our setting: the book *ETR*, the cash ETR (*CETR*) and a composite tax avoidance factor (*TAX_factor*).

ETR = income tax expenses/pre-tax income[7] (Chen et al., 2010; Porcano, 1986)

CETR = cash income taxes paid/pre-tax income[8] (Chen et al., 2010)

$TAX\_factor = 0.430 \times ETR + 0.770 \times ETR2 + 0.922 \times CETR + 0.922 \times CETR2 - 0.588 \times BTD - 0.588 \times DDBTD$ (Lennox et al., 2013)

Consistent with Tang et al. (2017) and Bradshaw et al. (2019), we focus on *ETR* and *CETR* given the validity problems besetting other measures in China. For example, prior studies find that book-tax difference measures are less informative of tax noncompliance in China where book-tax conformity is poor (Chan et al., 2010; Li et al., 2017).

Our third proxy is a composite measure of six indicators using factor analysis, *TAX_factor* (Lennox et al., 2013). In robustness tests, we also consider other tax measures, including the 3-year book ETR, the 3-year cash ETR, the differ-

[4] We follow prior research in reducing the impact of outliers by truncating ETR observations outside the [0,1] range (e.g., Chen et al., 2010; Law & Mills, 2015). We also verify that our core results hold when we winsorize ETR variables at the 0 and 1 values (e.g., Dyreng et al., 2008; McGuire et al., 2012).

[5] The estimation is likely to be biased if we include these political centers because some unobserved regional factors may simultaneously affect both the location of HSR lines and firms' tax avoidance decisions. Accordingly, we follow prior research by applying an inconsequential units approach by focusing on a sample of non-node cities whose characteristics are inconsequential to the choice of HSR lines (Faber, 2014; Redding & Turner, 2015; Cui & Li, 2019). In another robustness check, we also exclude those cities with a population over 5 million and find that the main results remain qualitatively similar.

[6] In a robustness check, we find that the main results are nearly identical when we exclude firms listed on the second board market (Growth Enterprise Market).

[7] In sensitivity analysis, we verify that our main results are materially insensitive to respecifying *ETR* with two alternative approaches: *ETR2* = income tax expense/(pre-tax income − deferred tax expense/statutory tax rate) (Shevlin, 1987); and *ETR3* = (income tax expense − deferred tax expense)/(pre-tax income − deferred tax expense/statutory tax rate) (Plesko, 2003).

[8] In sensitivity analysis, we verify that our core evidence holds after replacing *CETR* with two alternative specifications: *CETR2* = (cash income taxes paid)/(pre-tax income − deferred tax expense/statutory tax rate) and *CETR3* = (cash income taxes paid − Statutory tax rate × pre-tax income)/market value of the firm (Henry & Sansing, 2018). This measure benefits from recovering loss firms.

ence between the book and tax income (*BTD*) (Manzon & Plesko, 2001) and the residual of a regression that involves regressing BTD on total accruals (*DDBTD*) (Desai & Dharmapala, 2006).

3.2.3 | Measures for insiders' rent-extraction propensity

We specify several proxies for insiders' rent-extraction propensity. First, extensive prior research implies that insiders exploit the opacity surrounding *RPTS* to divert corporate resources at the expense of outside investors (Bae et al., 2002; Cheung et al., 2006; Jiang et al., 2010). It follows that the insiders' diversionary practices rise with the level of *RPT*. Second, given that firms usually have high other receivables when they prop earnings through *RPT*s, other receivables, *ORECTA*, can also reflect insiders' propensity to extract rents (Jian & Wong, 2010; Jiang et al., 2010). Third, dominant controlling shareholders routinely utilize the divergence between their control and cash flow rights (*DIVERGENCE*) to expropriate listed companies' resources (Claessens et al., 2002; Jiang et al., 2010). Fourth, given that rent-extraction opportunities rise when insiders are entrenched in firms with highly concentrated ownership structures, we also include the Herfindahl index of shareholdings of the 10 largest shareholders, *HHI10*, as one of the proxies for rent-extraction propensity. Finally, we also expect good corporate governance to alleviate agency costs and rent-extraction opportunities (Chen et al., 2012). Consequently, we analyze several corporate governance characteristics that reflect rent-extraction propensity, including the busyness of independent directors (*BUSYDIRECT*), the equity stake that the CEO (*CEO_OWN*) holds and the presence of a Big 10 auditor (*BIG10*) (Jian & Wong, 2010; Lennox et al., 2016).

3.2.4 | The HSR setting and regression models

The opening of a HSR line constitutes a shock that is external to firms' and local governments' decisions for two reasons. First, the Railway Plan was devised by the central government in striving to connect all major cities with a railway network. Accordingly, the decisions about when and where to launch HSR services are almost certainly insensitive to firm-level characteristics. Second, for cost efficiency and convenience, major cities are usually linked with straight HSR lines. As a result, most of the small cities in this network are included simply because they are located close to those straight lines, whereas their other characteristics are irrelevant. In summary, different cities are connected to the HSR system at different times for reasons beyond firms' and local governments' control. This alleviates endogeneity threats to reliable inference, enabling us to improve identification by gaging the impact of the launch of HSR lines in a staggered DID framework. To evaluate our presumption that the advent of HSR services encourages corporate site visits by external stakeholders, we estimate this model:

$$Visit_{it} = \beta_0 + \beta_1 HSR_{it} + CONTROLS_{it} + MACRO_{c,2003} \times T + \nu_i + \omega_t + \delta_p \times T + \eta_k \times T + \varepsilon_{it} \quad (1)$$

where *i*, *t*, *p*, c and *k* denote firm, year, province, city and industry, respectively. The dependent variable, $Visit_{it}$, represents the intensity of corporate site visits by institutional investors, financial analysts, individuals and news media. The key explanatory variable, HSR_{it}, is an indicator variable set to one for year *t* and onward if a HSR service is introduced during year *t* in the city in which a firm is headquartered and zero otherwise. Accordingly, the coefficient β_1 captures the effect of the launch of HSR services on the intensity of corporate site visits. We control for firm characteristics and corporate governance factors ($CONTROLS_{it}$) that may influence the incentives of external stakeholders to undertake corporate site visits according to extant research. These include firm size (*SIZE*), leverage (*LEV*), the return on equity (*ROE*), Tobin's *Q* (*Q*), loss status (*LOSS*), board size (*BSIZE*), the percentage of independent directors on the board (*BIND*), a CEO duality indicator (*DUAL*), the equity stake held by the largest shareholder (*TOP1*) and whether the government is the firm's ultimate controlling shareholder (*SOE*) (e.g., Chen, Ma, et al., 2022; Chen, Qu, et al., 2022;

Gul et al., 2010; Hutton et al., 2009). Additionally, we follow Wang (2013) by including city-specific trends in HSR line openings (*TREND*) to control for unobserved economic factors that are correlated with the timing of HSR construction. It is defined as $TREND = t - s_i$ if $t \geq s_i$ and 0 otherwise, where s_i denotes the year in which an HSR line opens in firm *i*'s city. To control for potential confounding industry and province-level effects (Gormley & Matsa, 2014), we add industry-by-year trend ($\eta_k \times T$) and province-by-year trend ($\delta_p \times T$) to the estimations.

The unbiased estimation of β_1 requires that the parallel trends assumption hold, which could be threatened by the non-random selection of the HSR routine. To alleviate this concern, we follow an approach used by Gentzkow (2006) and Li et al. (2020). This involves initially listing regional-level factors identified by existing studies as variables correlated with the choice of HSR location (Lin et al., 2023). These factors include the natural logarithm of a region's gross domestic product (*LNGDP*), the natural logarithm of the total population (*LNPEOPLE*), the natural logarithm of foreign direct investment inflows (*LNFDI*), the difference between fiscal expenditures and fiscal revenue scaled by the fiscal revenue (*DEFICIT*), the unemployment rate of the city (*UNEM*), the natural logarithm of passenger volume transported by conventional railways (*LNTRSP1*) and the natural logarithm of freight volume transported by conventional railways (*LNTRSP2*).[9] We include the year 2003 values of these variables ($MACRO_{2003}$) in the regressions.[10] Afterward, we interact the selected regional-level variables with the time trend ($MACRO_{c,2003} \times T$).[11]

We analyze the role that the launch of HSR services plays in shaping corporate tax avoidance with the following model:

$$TAX_{it} = \beta_0 + \beta_1 HSR_{it} + CONTROLS_{it} + MACRO_{c,2003} \times T + \nu_i + \omega_t + \delta_p \times T + \eta_k \times T + \varepsilon_{it} \quad (2)$$

where *TAX* is the tax avoidance measure. We expect to observe under the prediction in H1 that β_1 is significantly positive. Grounded in prior work on the determinants of tax avoidance in China (Li et al., 2017; Lin et al., 2018; Zhang et al., 2016), we control for firm size (*SIZE*), leverage (*LEV*), the *ROE*,[12] Tobin's *Q* (*Q*), fixed assets (*PPE*), intangible assets (*INTANG*), inventory intensity (*INVENTORY*), *LOSS* status and SOE status.

In all regressions, we also control for firm and year fixed effects, province-by-year trends and industry-by-year trends. We winsorize all continuous variables at the 1st and 99th percentiles. Standard errors are clustered at the firm level.[13] We provide detailed variable definitions in Appendix A.

3.3 | Descriptive statistics

Table 1 reports the distribution of our sample across each industry for each year. Most of the observations are from the manufacturing industry, accounting for almost 79% of the sample, which partly reflects that we exclude service firms when they are located in political centers.[14] Overall, there is an upward trend in the number of observations over time.

In Table 2, we report descriptive statistics for the main variables. The mean (median) values of *ETR*, *CETR* and *Tax_factor* are 0.151 (0.141), 0.179 (0.149) and −0.039 (−0.213), respectively, which are generally comparable with

[9] In a separate test, we verify that these macroeconomic variables are closely associated with the timing of a city's HSR launch. These results are detailed in the Supporting Information Appendix A1. To mitigate potential endogeneity concerns, we have incorporated these variables into our regression analysis.

[10] The HSR system was first proposed in the national medium and long-term railway network plan, which was released by the State Council of China in 2004. As such, we use the year 2003 value of these variables.

[11] Consistent with Gentzkow (2006) and Li et al. (2020), we control for the interactions between the 2003 values of the macroeconomic variables and the time trend *T*, instead of the annual values of these variables. This design choice reflects that the HSR opening is likely to influence the values of these macroeconomic variables, potentially leading to an endogeneity issue.

[12] We continue to find supportive evidence when we replace *ROE* with return on assets.

[13] The core results remain qualitatively identical when we cluster the standard errors at the city level.

[14] In sensitivity analysis, we verify that our main results are nearly identical when we restrict the sample to only firms belonging to the manufacturing industry. We also continue to find supportive evidence when we exclude firms cross-listed in foreign markets from the sample.

TABLE 1 Sample distribution by year and industry.

Industry	Year																		
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
A	7	8	4	9	7	14	16	10	10	11	12	15	8	10	8	12	11	10	182
B	7	11	15	16	16	21	22	25	25	23	11	15	19	19	21	19	22	23	330
C	314	362	393	366	432	608	731	752	731	743	759	902	1098	1047	1079	1224	1306	1205	14,052
D	15	13	14	11	17	16	16	19	18	13	17	17	17	15	18	19	23	19	297
E	6	8	8	6	9	11	15	14	14	15	20	28	31	24	25	19	18	17	288
F	18	21	21	24	24	31	36	35	38	33	34	37	42	38	42	43	43	40	600
G	15	19	20	18	22	25	26	30	29	28	30	28	29	32	32	30	31	32	476
H	2	0	1	2	0	2	1	3	3	2	1	1	1	1	1	0	0	0	21
I	3	5	5	6	6	15	20	19	17	20	21	33	42	34	40	45	45	32	408
K	13	15	18	16	22	27	27	34	33	35	28	30	25	24	23	24	22	17	433
L	6	5	6	7	6	9	9	10	9	10	11	16	14	11	16	11	13	9	178
M	0	0	0	0	1	2	3	4	4	6	6	8	12	15	14	10	16	15	116
N	1	1	1	2	1	2	2	9	10	12	12	13	18	23	22	19	23	18	189
O	0	0	0	2	3	2	3	0	0	0	0	0	0	0	0	0	0	0	10
P	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	2	5
Q	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	1	0	0	5
R	0	0	0	1	1	1	1	1	2	1	4	4	7	3	4	2	4	2	38
S	17	17	20	20	20	16	21	7	9	9	6	6	7	6	9	6	6	4	206
Total	424	485	526	506	587	802	949	972	952	961	972	1154	1372	1304	1355	1485	1583	1445	17,834

Note: This table presents the distribution of our sample. A: agriculture, forestry, husbandry and fishery; B: mining; C: manufacturing; D: electric power and heat production and supply; E: construction; F: wholesale and retail; G: transport, storage and postal service; H: accommodation and catering; I: software and information technology services; K: real estate; L: leasing and commercial service; M: scientific research and technical service; N: water conservancy, environment and public facility; O: resident service, repair and other services; P: education; Q: health and social work; R: culture, sports and entertainment and S: diversified industry.

TABLE 2 Descriptive statistics.

Variable	*N*	Mean	SD	Median	Min	Max
ETR	17,834	0.151	0.085	0.141	0	0.483
CETR	15,982	0.179	0.124	0.149	0.003	0.710
TAX_factor	14,421	−0.039	0.926	−0.213	−1.996	3.260
Visit_Inst	17,834	0.995	1.479	0	0	4.564
Visit_Anlyst	17,834	0.885	1.294	0	0	3.850
Visit_Indiv	17,834	0.041	0.163	0	0	0.693
Visit_Media	17,834	0.036	0.153	0	0	0.693
HSR	17,834	0.747	0.434	1	0	1
SIZE	17,834	21.94	1.188	21.79	19.29	25.46
LEV	17,834	0.399	0.193	0.392	0.072	0.885
ROE	17,834	0.093	0.067	0.081	0.001	0.390
Q	17,834	1.961	1.114	1.585	0.917	6.507
PPE	17,834	0.230	0.151	0.204	0.004	0.681
INTANG	17,834	0.04	0.044	0.034	0	0.273
INVENTORY	17,834	0.145	0.116	0.120	0	0.646
LOSS	17,834	0.0610	0.239	0	0	1
SOE	17,834	0.292	0.455	0	0	1
BSIZE	17,834	8.567	1.695	9	5	15
BIND	17,834	37.30	5.159	33.33	30	57.14
DUAL	17,834	0.316	0.465	0	0	1
TOP1	17,677	34.40	14.32	32.32	9.030	73.03
LNGDP	17,834	6.818	0.891	6.903	4.635	7.971
LNPEO	17,834	5.849	0.623	6.071	4.407	6.982
LNFDI	17,834	11.48	1.779	11.92	7.120	14.04
DEFI	17,834	0.393	0.404	0.223	0.034	2.096
UNEM	17,834	0.052	0.024	0.056	0.010	0.138
LNTRSP1	17,834	9.000	0.745	9.305	7.148	10.39
LNTRSP2	17,834	8.687	0.615	8.824	6.452	10.33

Note: This table presents some descriptive statistics. The sample size for variables used in the main regressions reported in Table 4 is fixed at 17,834. Definitions of all variables are presented in Appendix A.

those in prior studies covering similar timeframes (e.g., Bauer et al., 2020; Lin et al., 2018). Overall, 74.7% of the observations fall in the post-HSR period, reflecting the growing prevalence of HSR services in China over time.

4 | EMPIRICAL RESULTS

4.1 | The launch of HSR services and corporate site visits

In this section, we evaluate whether the launch of HSR services induces corporate site visits by external stakeholders. Consistent with this proposition, the regression results reported in Table 3 imply that the intensity of site visits that

TABLE 3 High-speed rail (HSR) and site visits that are undertaken by external stakeholders.

	(1)	(2)	(3)	(4)
	Visit_Inst	*Visit_Anlyst*	*Visit_Indiv*	*Visit_Media*
HSR	0.100**	0.092**	−0.004	−0.005
	(2.569)	(2.564)	(−0.766)	(−0.832)
SIZE	0.573***	0.506***	0.012***	0.016***
	(17.794)	(18.546)	(3.615)	(4.602)
LEV	−0.178*	−0.196**	−0.014	−0.007
	(−1.880)	(−2.357)	(−1.133)	(−0.651)
ROE	0.325***	0.293***	−0.004	0.005
	(7.741)	(7.777)	(−0.747)	(0.923)
Q	0.176***	0.147***	0.004**	0.004***
	(12.684)	(12.531)	(2.361)	(2.609)
LOSS	−0.142***	−0.162***	−0.002	0.001
	(−4.969)	(−6.901)	(−0.539)	(0.354)
BSIZE	−0.020*	−0.014	−0.000	−0.000
	(−1.666)	(−1.361)	(−0.309)	(−0.227)
BIND	−0.005	−0.004	−0.000	0.000
	(−1.589)	(−1.589)	(−0.290)	(0.146)
DUAL	0.018	0.026	−0.007	0.002
	(0.537)	(0.915)	(−1.639)	(0.542)
TOP1	−0.004*	−0.001	−0.000	0.000
	(−1.935)	(−0.302)	(−0.914)	(1.344)
SOE	−0.127*	−0.157***	0.000	−0.000
	(−1.833)	(−2.624)	(0.048)	(−0.033)
TREND	0.004	−0.007	−0.002	−0.003*
	(0.395)	(−0.734)	(−1.332)	(−1.943)
Firm and year FE	Yes	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes	Yes
N	20,279	20,279	20,279	20,279
Adj R^2	0.550	0.549	0.124	0.104
F	28.55	30.27	1.941	1.974

Note: This table presents the results of examining the impact of HSR line inaugurations on site visits by external stakeholders, estimated from the following OLS equation: $Visit_{it} = \beta_0 + \beta_1 HSR_{it} + CONTROLS_{it} + MACRO_{c,2003} \times T + \nu_i + \omega_t + \delta_p \times T + \eta_k \times T + \varepsilon_{it}$, where *i*, *t*, *p*, c and *k* denote firm, year, province, city and industry, respectively. $Visit_{it}$ represents the intensity of corporate site visits by institutional investors (*Visit_Inst*), financial analysts (*Visit_Anlyst*), individuals (*Visit_Indiv*) and news media (*Visit_Media*). HSR_{it} is an indicator variable set to one for year *t* and onward if an HSR service is introduced during year *t* in the city in which a firm is headquartered and zero otherwise. The *CONTROLS* includes *SIZE*, *LEV*, *ROE*, *Q*, *LOSS*, *BSIZE*, *BIND*, *DUAL*, *TOP1* and *SOE*. *MACRO* includes *LNGDP*, *LNPEOPLE*, *LNFDI*, *DEFICIT*, *UNEM*, *LNTRSP1* and *LNTRSP2*. All models include firm and year fixed effect, industry-by-year trend and province-by-year trend. We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.

***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

are undertaken by institutional investors and financial analysts rises. However, other results indicate that individual investors and news media do not increase their site visits after the initiation of HSR services. Given that we observe no change in their information acquisition, we exclude site visits involving these stakeholders from subsequent tests.

4.2 | Main results

4.2.1 | The launch of HSR services and tax avoidance

Next, we examine our first hypothesis that the launch of HSR services leads to firms taking less aggressive tax positions. In Panel A of Table 4, we report that the coefficients on *HSR* are positive and statistically significant at the 10% level (two-tailed) or better for all three tax avoidance proxies (coeff = 0.011, 0.010 and 0.082; *t* stats = 3.059, 1.913 and 2.314). Lending support to the prediction in H1, this evidence implies that tax avoidance falls after HSR services become available. Reflecting the first-order economic impact according to the coefficient estimates, tax avoidance declines, for example, by 1.10% in absolute magnitude and 7.28% in relative terms (0.011/0.151) for *ETR* in Column 1; the economic importance is also material for *CETR* and *Tax_factor*. The results for the control variables are generally consistent with prior research (e.g., Li et al., 2017; Chen, Tang, et al., 2021).

Our sample covers the period from 2005 to 2022. However, the latter years in this timeframe—specifically, from 2020 to 2022—were majorly affected by the Covid-19 pandemic that caused major disruptions in railway operations and largely restricted people's travel. Consequently, we anticipate a weaker effect of HSR line introductions on tax avoidance during this period. We designate 2020–2022 as the Covid-19-impacted period, given that the pandemic started affecting operations at the end of 2019 and most travel restrictions remained in place until the end of 2022. In Panel B, we report the results from reestimating the main regression after isolating this subperiod. We find that the HSR coefficients are considerably smaller than those for the entire sample period and are no longer statistically significant. This outcome aligns with our expectations and lends support to the validity of our identification.

Moreover, if the launch of HSR services indeed curtails firms' overinvestment in tax avoidance, we would expect this effect to be more concentrated among firms that were more likely to overinvest in tax avoidance before the introduction of HSR. To validate this conjecture, we follow prior research by employing a quantile regression approach to explore the relationship between tax avoidance and HSR services across different quantiles of the tax avoidance distribution (e.g., Armstrong et al., 2015; Hoopes et al., 2012). We anticipate that firms with lower ETR values, suggesting heavier investments in tax avoidance, exhibit this effect more strongly. Consistent with expectations, the results reported in Panel C imply that HSR plays a larger role at the lower percentiles of ETR for all three ETR measures, diminishing monotonically toward the higher percentiles. This pattern reflects that HSR services play a major role in dampening firms' excessive tax avoidance efforts.

4.2.2 | The effect of corporate site visits

To further identify which stakeholders are behind the reduced tax avoidance after the opening of HSR lines, we analyze the impact of the increased intensity of site visits stemming from the launch of HSR lines on tax avoidance. Consistent with the two-step method in Core et al. (1999) and Mao (2021), we compute the increase in the intensity of site visits as follows:

$$Pre\_Visit_{i,t} = \hat{\beta}_1 HSR_{i,t} \tag{3}$$

where the estimated coefficients on the *HSR* variables ($\hat{\beta}_1$) are those reported in Table 3. It represents the increased intensity of site visits arising from the opening of HSR lines, which considers all other known determinants. As shown

TABLE 4 High-speed rail (HSR) and tax avoidance: baseline regression results.

Panel A: The full sample			
	(1)	(2)	(3)
	ETR	*CETR*	*TAX_factor*
HSR	0.011***	0.010*	0.082**
	(3.059)	(1.913)	(2.314)
SIZE	−0.004*	−0.006	−0.011
	(−1.663)	(−1.412)	(−0.371)
LEV	−0.062***	0.062***	0.949***
	(−6.860)	(3.921)	(7.215)
ROE	0.023**	−0.319***	−4.361***
	(2.351)	(−10.601)	(−10.851)
Q	−0.004***	−0.006***	−0.059***
	(−4.084)	(−3.843)	(−4.029)
PPE	−0.046***	0.015	0.139
	(−4.444)	(0.920)	(1.097)
INTANG	−0.079**	0.005	0.112
	(−2.439)	(0.105)	(0.305)
INVENTORY	0.041**	0.057**	0.443**
	(2.350)	(2.261)	(2.395)
LOSS	−0.015***	−0.009	−0.150***
	(−3.901)	(−1.208)	(−3.055)
SOE	0.004	−0.002	0.032
	(0.590)	(−0.216)	(0.515)
TREND	0.000	0.001	0.013
	(0.088)	(0.605)	(1.300)
Firm and year FE	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes
N	17,834	16,173	14,208
Adj R^2	0.405	0.250	0.460
F	12.40	11.99	25.32
Panel B: The Covid-19 pandemic subperiod			
	(1)	(2)	(3)
	ETR	*CETR*	*TAX_factor*
HSR	0.002	−0.011	−0.014
	(0.069)	(−0.484)	(−0.071)
SIZE	−0.024**	−0.035**	−0.388***
	(−2.355)	(−2.321)	(−4.422)

(Continues)

TABLE 4 (Continued)

Panel B: The Covid-19 pandemic subperiod			
	(1)	(2)	(3)
	ETR	*CETR*	*TAX_factor*
LEV	−0.035	−0.031	0.453
	(−1.365)	(−0.696)	(1.523)
ROE	0.013	−0.416***	−5.479***
	(0.458)	(−8.023)	(−10.625)
Q	−0.001	−0.002	−0.043*
	(−0.614)	(−0.658)	(−1.906)
PPE	−0.024	−0.000	0.210
	(−0.751)	(−0.011)	(0.837)
INTANG	−0.043	0.055	0.632
	(−0.451)	(0.398)	(0.662)
INVENTORY	−0.030	−0.004	0.728
	(−0.641)	(−0.044)	(1.554)
LOSS	0.015	−0.001	0.066
	(1.200)	(−0.032)	(0.656)
SOE	0.004	−0.040	−0.207*
	(0.182)	(−1.409)	(−1.871)
TREND	−0.004	−0.006	−0.029
	(−1.257)	(−0.900)	(−0.607)
Firm and year FE	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes
N	4277	3899	3459
Adj R^2	0.540	0.424	0.690
F	3.701	19.73	29.97

Panel C: Quantile regression analysis									
	ETR as the dependent variable								
Percentile	**10th**	**20th**	**30th**	**40th**	**50th**	**60th**	**70th**	**80th**	**90th**
HSR	0.014***	0.013***	0.012***	0.011***	0.011***	0.010***	0.009**	0.008*	0.007
	(4.297)	(4.216)	(4.030)	(3.742)	(3.334)	(2.864)	(2.376)	(1.922)	(1.403)
CETR as the dependent variable									
HSR	0.012***	0.012***	0.011**	0.011**	0.010**	0.010*	0.009	0.008	0.007
	(2.650)	(2.666)	(2.551)	(2.360)	(2.077)	(1.725)	(1.373)	(1.054)	(0.733)
TAX_factor as the dependent variable									
HSR	0.122**	0.103**	0.090**	0.079**	0.069**	0.060*	0.052	0.044	0.031
	(2.085)	(2.179)	(2.209)	(2.178)	(2.059)	(1.870)	(1.622)	(1.329)	(0.844)

(Continues)

TABLE 4 (Continued)

Note: Panel A presents the results of OLS regressions that examine the impact of HSR line inaugurations on corporate tax avoidance over the whole sample period, spanning from 2005 to 2022, whereas Panel B presents the results during the Covid-19 pandemic subperiod, spanning from 2020 to 2022. We estimate the following equation: $TAX_{it} = \beta_0 + \beta_1 HSR_{it} + CONTROLS_{it} + MACRO_{c,2003} \times T + \nu_i + \omega_t + \delta_p \times T + \eta_k \times T + \varepsilon_{it}$, where *i*, *t*, *p*, *c* and *k* denote firm, year, province, city and industry, respectively. Columns 1–3 use *ETR* (income tax expenses/pre-tax income), *CETR* (cash income taxes paid/pre-tax income) and *TAX_factor* (the composite measure using factor analysis) to measure corporate tax avoidance. HSR_{it} is an indicator variable set to one for year *t* and onward if a HSR service is introduced during year *t* in the city in which a firm is headquartered, and zero otherwise. The *CONTROLS* includes *SIZE*, *LEV*, *ROE*, *Q*, *PPE*, *INTANG*, *INVENTORY*, *LOSS* and *SOE*. *MACRO* includes *LNGDP*, *LNPEOPLE*, *LNFDI*, *DEFICIT*, *UNEM*, *LNTRSP1* and *LNTRSP2*. All models include firm and year fixed effect, industry-by-year trend and province-by-year trend. We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Panel C presents coefficient estimates of *HSR* at various quantiles of the tax avoidance distribution (from 10th to 90th percentiles). We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.
***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

in Table 5, *Pre_Visit_Inst* and *Pre_Visit_Anlyst* are positively associated with all three ETR measures, implying that firms reduce their tax avoidance after the increase in the intensity of site visits conducted by institutional investors and financial analysts due to the HSR introduction. This finding suggests that the launch of HSR services facilitates institutional investors and financial analysts in acquiring firm-specific information and disciplining insiders' behavior.

4.3 | Robustness tests

To address the concern that the reduction in tax avoidance may spuriously stem from some confounding factors, we conduct several robustness tests.

4.3.1 | The dynamic effect of HSR line openings on tax avoidance

If the fall in firms' tax avoidance is driven by the launch of HSR services, we should find that the difference in tax avoidance between the treatment group (firms in cities that are connected by HSR lines) and the control group (firms in cities that are not connected by HSR lines) only manifests in the years after the opening of HSR lines, rather than in earlier years. To verify this, we follow Li et al. (2017) and Bai et al. (2020) in estimating this dynamic regression equation:

$$\begin{aligned} TAX_{i,t} = {} & \alpha + \beta_{-3} HSR_{i,t}^{-3} + \beta_{-2} HSR_{i,t}^{-2} + \beta_{-1} HSR_{i,t}^{-1} + \beta_0 HSR_{i,t}^{0} + \beta_1 HSR_{i,t}^{+1} + \beta_2 HSR_{i,t}^{+2} + \beta_3 HSR_{i,t}^{+3} \\ & + \beta_4 HSR_{i,t}^{\geq +4} + \gamma CONTROLS_{i,t} + MACRO_{c,2003} \times T + \nu_i + \omega_t + \delta_p \times T + \eta_k \times T + \varepsilon_{it} \end{aligned} \quad (4)$$

where $HSR_{i,t}^{j}$ is an indicator variable that equals one if a firm is headquartered in a city that has been connected by an HSR line for *j* years and zero otherwise. In Appendix A, we provide definitions for the rest of the variables.

As shown in Table 6, the coefficients on HSR^{-3}, HSR^{-2} and HSR^{-1} are statistically indistinguishable from zero, whereas the coefficients on HSR^{0}, HSR^{+1}, HSR^{+2}, HSR^{+3} and $HSR^{\geq +4}$ generally enter positively. This analysis supports that there is no perceptible difference in the tax avoidance level between the treatment and the control groups before HSR line openings, whereas the difference between the two groups becomes significant afterward.

TABLE 5 HSR, site visits and tax avoidance.

	(1)	(2)	(3)	(4)	(5)	(6)
	ETR	*CETR*	*TAX_FACTOR*	*ETR*	*CETR*	*TAX_factor*
Pre_Visit_Inst	0.084***	0.090*	0.695*			
	(3.043)	(1.688)	(1.878)			
Pre_Visit_Anlyst				0.092***	0.099*	0.761*
				(3.043)	(1.688)	(1.878)
SIZE	−0.002	−0.002	0.015	−0.002	−0.002	0.015
	(−1.026)	(−0.719)	(0.595)	(−1.026)	(−0.719)	(0.595)
LEV	−0.067***	0.048***	0.875***	−0.067***	0.048***	0.875***
	(−9.347)	(3.364)	(7.073)	(−9.347)	(3.364)	(7.073)
ROE	0.018*	−0.314***	−4.358***	0.018*	−0.314***	−4.358***
	(1.959)	(−12.143)	(−12.466)	(1.959)	(−12.143)	(−12.466)
Q	−0.004***	−0.006***	−0.056***	−0.004***	−0.006***	−0.056***
	(−4.406)	(−4.208)	(−4.291)	(−4.406)	(−4.208)	(−4.291)
PPE	−0.035***	0.024	0.247**	−0.035***	0.024	0.247**
	(−4.155)	(1.504)	(2.103)	(−4.155)	(1.504)	(2.103)
INTANG	−0.057**	0.018	0.270	−0.057**	0.018	0.270
	(−2.314)	(0.432)	(0.872)	(−2.314)	(0.432)	(0.872)
INVENTORY	0.035**	0.062**	0.457***	0.035**	0.062**	0.457***
	(2.488)	(2.487)	(2.639)	(2.488)	(2.487)	(2.639)
LOSS	−0.013***	−0.004	−0.139***	−0.013***	−0.004	−0.139***
	(−3.440)	(−0.509)	(−2.849)	(−3.440)	(−0.509)	(−2.849)
SOE	0.001	−0.005	0.045	0.001	−0.005	0.045
	(0.197)	(−0.612)	(0.836)	(0.197)	(−0.612)	(0.836)
TREND	−0.002	−0.002	0.015	−0.002	−0.002	0.015
	(−1.026)	(−0.719)	(0.595)	(−1.026)	(−0.719)	(0.595)
Firm and year FE	Yes	Yes	Yes	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes	Yes	Yes	Yes
N	16,916	15,386	13,481	16,916	15,386	13,481
Adj R^2	0.418	0.269	0.485	0.418	0.269	0.485
F	15.84	16.74	35.27	15.84	16.74	35.27

Note: This table presents the results of OLS regressions that examine the impact of increased site visits by institutional investors and financial analysts, attributable to the opening of HSR lines, on tax avoidance. *Pre_Visit_Inst and Pre_Visit_Anlyst*, which are proxies for the increase in the intensity of site visits due to the opening of HSR lines, are estimated with Equation (3). We present *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.

***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

TABLE 6 Dynamic effects of high-speed rail (HSR) on tax avoidance—dynamic analysis.

	(1) ETR	(2) CETR	(3) TAX_factor
HSR^{-3}	−0.001	0.010	0.006
	(−0.164)	(1.296)	(0.095)
HSR^{-2}	0.010	0.012	0.051
	(1.598)	(1.393)	(0.708)
HSR^{-1}	0.007	0.014	0.101
	(1.051)	(1.300)	(1.199)
HSR^{0}	0.014*	0.024**	0.181**
	(1.672)	(2.072)	(2.026)
HSR^{+1}	0.019**	0.026**	0.206**
	(2.103)	(1.992)	(2.119)
HSR^{+2}	0.015*	0.019	0.177*
	(1.653)	(1.432)	(1.820)
HSR^{+3}	0.016*	0.018	0.158
	(1.778)	(1.411)	(1.608)
$HSR^{\geq +4}$	0.017*	0.020	0.143
	(1.716)	(1.486)	(1.402)
Control	Yes	Yes	Yes
Firm and year FE	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes
N	17,760	16,123	14,132
Adj R^2	0.430	0.265	0.485
F	8.064	9.002	19.20

Note: This table presents the results of OLS regressions on the dynamic effect of HSR line openings on corporate tax avoidance. $HSR^{j}_{i,t}$ is an indicator variable that equals one if a firm is headquartered in a city connected by an HSR line *j* years before and zero otherwise. The control variables are the same as specified in Table 4. We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.
***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

In short, this evidence collectively suggests that the introduction of HSR is responsible for the shift in firms' tax avoidance.[15]

4.3.2 | Placebo tests

To further explore the competing explanation that our results spuriously stem from some unobservable shocks that coincide with the HSR events, we follow recent research by conducting placebo tests based on artificial events (e.g.,

[15] A reverse causality problem may ensue if some firms elect to relocate to be closer to the HSR before its launch. However, we verify that our core evidence remains after we exclude relocating firms.

Cornaggia et al., 2015; Li et al., 2020).[16] This involves counting the actual number of cities that are newly connected by HSR lines in each year and then randomly assigning the same number of cities (without replacement) to a pseudo-HSR-connection group. This approach maintains the distribution of HSR connection across years from our baseline specification, while disrupting the proper assignment of HSR connection to cities. Given the random data generation process, pseudo-HSR-connection should be irrelevant to tax avoidance; otherwise, it would suggest that our analysis suffers from an omitted variable problem of unknown severity. To yield a rigorous estimation of coefficients in Equation (2), we repeat the process 1000 times.

Figure 1 reports the histogram of these estimated coefficients. To facilitate comparisons, we also add vertical lines that represent the actual coefficients on *HSR* from Table 4. This analysis implies that irrespective of whether we specify *ETR*, *CETR* or *Tax_factor* to measure tax avoidance, the vast majority of coefficients derived from the random pseudo-HSR-connection cluster around zero, and no more than 5% of them are larger than the true coefficients. These results further dispel the concern that unobserved factors are behind our core evidence.

4.3.3 | Stacked DID with entropy balancing

Recent research by Goodman-Bacon (2021) and Baker et al. (2022) highlights challenges besetting the traditional staggered DID method. These challenges stem primarily from time-related fluctuations and variations in treatment effects. Accordingly, we follow Cengiz et al. (2019) and Baker et al. (2022) by employing a stacked DID approach.

In applying this approach, we compile a dataset comprising both treated and control firms for each HSR service launch event. Within this dataset, firms that receive their first HSR connection in a designated year, labeled as t, belong to the treated category. In contrast, the control group comprises firms without an HSR connection from the onset of our sample period up to 4 years past the event, that is, until $t + 4$.

To ensure a fair comparison between the treated and control groups, we implement entropy balancing,[17] taking into account control variables from the year just before the event ($t - 1$). This strategy aims to account for inherent differences between the two groups. In considering a 6-year period (spanning 3 years before and 3 years after the event) for each firm, we group our data into separate cohorts. After stacking these cohorts, we derive a new dataset for our regression analysis.

In Panel A of Table 7, we observe that the treatment firms significantly differ from the control firms in several ways. However, after applying entropy balancing, no significant differences remain in the firm-level characteristics, suggesting that we reach covariate balance.

Next, we run the following regression:

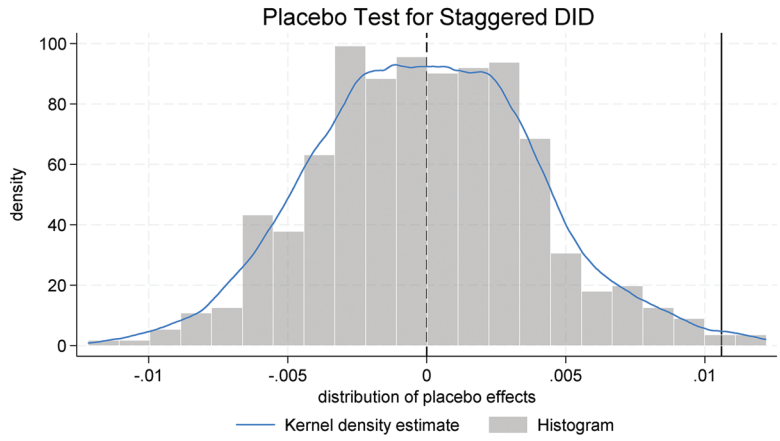
$$\begin{aligned} TAX_{i,h,t} = \ & \alpha + \beta_1 Treatment_{i,h} \times Post_{h,t} + \gamma CONTROLS_{i,h,t} + MACRO_{c,2003} \times T \\ & + \nu_{i,h} + \omega_{h,t} + \delta_p \times T + \eta_k \times T + \varepsilon_{it} \end{aligned} \tag{5}$$

where i, h and t indicate firm, cohort and year, respectively. *Treatment* is a dummy variable assigned a value of one for firms in the treatment group and zero for others. *Post* is another dummy variable taking the value of one for fiscal years after the initiation of HSR services and zero in other cases. The interaction term, *Treatment* × *Post*, represents the impact of HSR connections on tax avoidance.

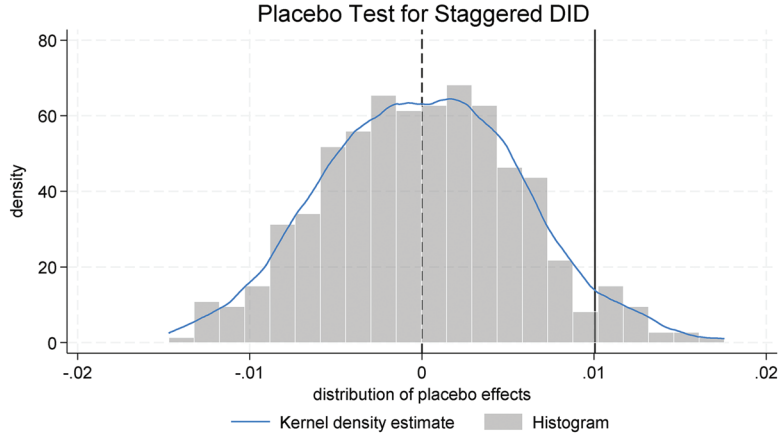
[16] In another form of placebo tests, we verify that we no longer find supportive evidence when we randomize the event years. This involves randomly selecting two years prior to the actual opening years of the HSR lines as the opening years and re-estimating Equation (2). Consistent with expectations, we find that firms' tax avoidance is insensitive to the falsified opening of HSR lines. We obtain similar results when we take four or five years prior to the actual HSR launch in falsifying these years.

[17] Additionally, we applied propensity score matching with replacement using caliper widths ranging from 0.01 to 0.1 and setting different matching ratios (i.e., 1:1, 1:2 and 1:3). Our core results hold in all these estimations.

Panel (A): *ETR* as the proxy for tax avoidance (*p* value=0.006)



Panel (B): *CETR* as the proxy for tax avoidance (*p* value=0.048)



Panel (C): *TAX_factor* as the proxy for tax avoidance (*p* value=0.026)

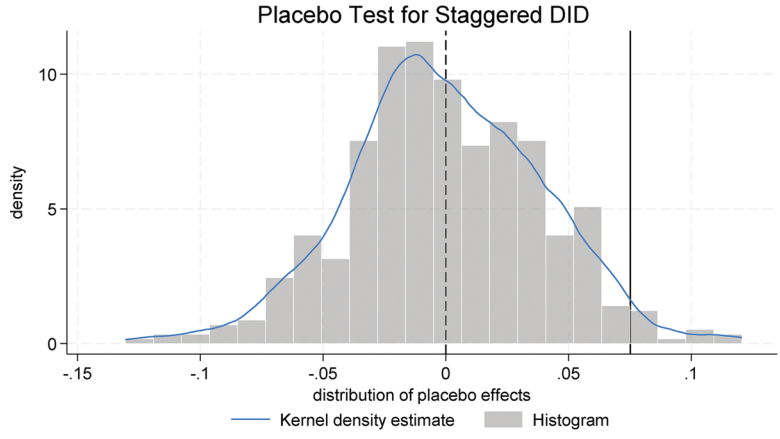


FIGURE 1 Distribution of estimated coefficients of falsification tests. *Notes*: These figures show the distribution of the estimated coefficients of high-speed rail (*HSR*) from 1000 simulations that randomly assign some city-year observations to a pseudo-HSR-connection group. Panel (A) to Panel (C) use *ETR*, *CETR*, and *TAX_factor* as the proxy for tax avoidance, respectively. The vertical dash lines represent the coefficients of *HSR* from the baseline regression.

TABLE 7 High-speed rail (HSR) and tax avoidance: stacked difference-in-differences (DID) with entropy balancing approach.

Panel A: Comparison of means before and after entropy balancing						
	Before entropy balancing			After entropy balancing		
	Treatment ($n = 1463$)	Control ($n = 2244$)	*Diff*	Treatment ($n = 1463$)	Control ($n = 2244$)	*Diff*
ETR	0.158	0.149	−0.009***	0.158	0.158	0.000
SIZE	21.801	21.745	−0.056	21.801	21.801	0.000
LEV	0.444	0.450	0.007	0.444	0.444	0.000
ROE	0.092	0.096	0.005	0.092	0.092	0.000
Q	2.007	2.063	0.056	2.007	2.007	0.000
PPE	0.242	0.292	0.050***	0.242	0.242	0.000
INTANG	0.046	0.050	0.004**	0.046	0.046	0.000
INVENTORY	0.167	0.159	−0.008	0.167	0.167	0.000
LOSS	0.070	0.096	0.025***	0.070	0.070	0.000
SOE	0.485	0.558	0.073***	0.485	0.485	0.000

Panel B: Regression results using entropy-balanced samples						
	ETR		*CETR*		*TAX_factor*	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment × *Post*	0.014*	0.013*	0.030**	0.027**	0.186	0.170*
	(1.875)	(1.781)	(2.098)	(2.055)	(1.605)	(1.719)
Control	No	Yes	No	Yes	No	Yes
Firm-Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
$MACRO_{2003} \times T$	No	Yes	No	Yes	No	Yes
Industry × *T*	Yes	Yes	Yes	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes	Yes	Yes	Yes
N	19,156	18,662	17,867	17,459	16,320	16,017
Adj R^2	0.545	0.564	0.410	0.479	0.497	0.633
F	2.097	6.062	1.642	7.236	1.168	57.95

Notes: This table presents the results obtained using the stacked DID approach, complemented by entropy balancing. Panel A compares means along various sample dimensions before and after entropy balancing. Panel B reports the regression results. All variables are defined in Appendix A. The *t*-statistics based on robust standard errors clustered at the firm level are in parentheses.

*, ** and *** indicate statistical significance for a two-tailed test at the 10%, 5% and 1% levels, respectively.

In Panel B of Table 7, the coefficients on *Treatment* × *Post* are positive and statistically significant in five of the six columns that span various regression specifications, reinforcing our earlier evidence implying that the launch of HSR services curtails tax avoidance.

4.3.4 | Instrumental variable (IV) approach

In another standard technique for tackling endogeneity, we utilize an IV approach. Prior research suggests that historical transportation infrastructure data are a valid instrument given that they have a strong connection to the evolution

of current transportation systems yet do not stem from current business activities or local governmental policies (e.g., Baum-Snow et al., 2017; Duranton & Turner, 2012; Moller & Zierer, 2018). We follow Baum-Snow et al. (2017), Dong et al. (2020) and Chen, Chi, et al. (2021) by specifying an indicator variable, $RAILWAY_{c,\,1962}$, to identify whether a city c had a railway station back in 1962. These studies outline several reasons for selecting this instrument. First, the railways built before 1962 were largely under Soviet influence to connect resource-rich regions with manufacturing centers. Since then, China's economy has gradually shifted from an agriculture-oriented, Soviet-style model to an industrialized, export- and market-oriented model. Accordingly, the railways built before 1962 are unlikely to have any relation to present-day corporate tax decisions. Moreover, these early rail lines essentially laid the groundwork for the country's current railway systems and are directly relevant to the evolution of modern HSR lines. As such, this instrument variable satisfies both the exclusion restriction and the relevance condition.

However, because $RAILWAY_{c,\,1962}$ remains constant over time, it is not suitable for the IV analysis on its own. To account for time-varying effects, we interact them with the year-fixed effects (e.g., Duflo & Pande, 2007).

In this analysis, we rely on a two-stage least squares regression. In the first stage, we regress the initiation of HSR services on the interaction between $RAILWAY_{c,\,1962}$ and *Year dummies*, along with other control variables. In the second stage, we regress our tax avoidance measures on the predicted value of HSR derived from the first stage, alongside other controls.

In untabulated first-stage results, we find that most of the coefficient estimates for the interaction between $RAILWAY_{c,\,1962}$ and *Year dummies* are significantly positive and the first-stage regression *F* values are larger than the rule-of-thumb value of 10, attesting to the instrument's validity. In the results of the second stage regression reported in Table 8, the coefficients on the predicted value of HSR enter positively in all cases, corroborating our earlier evidence on the importance of HSR to tax avoidance.

Additionally, we specify tax avoidance with the 3-year book ETR, the 3-year cash ETR, book-tax differences (Manzon & Plesko, 2001) and residual book-tax differences (Desai & Dharmapala, 2006) in successive regressions and find that our baseline results hold.[18]

4.4 | The effect of agency costs

The evidence so far suggests that external stakeholders' information acquisition costs subside and firms undertake less aggressive tax positions after the launch of HSR services. In this section, we deepen the analysis by exploring why the initiation of HSR services affects tax avoidance.

4.4.1 | The effect of insiders' rent-extraction propensity

If the fall in tax avoidance after the opening of HSR lines stems from it becoming harder for insiders to extract private benefits, we should observe under the prediction in H2 that the effect is concentrated in firms in which insiders exhibit a high rent-extraction propensity. To examine this agency cost hypothesis, we bisect the full sample into paired subgroups based on the yearly median values of the proxies for rent-extraction propensity and reestimate Equation (2) for each sub-group. In Table 9, we report that the coefficients on *HSR* generally only enter positively when we isolate firms with higher rent-extraction propensity. For example, *ETR* increases by 1.6% in firms with high *RPT* levels after the launch of HSR lines, whereas the change in *ETR* is statistically indistinguishable from zero for their counterparts; the

[18] We also verify that our core results are almost identical when we exclude the top 5% of the firms in our sample according to total assets on the grounds that larger firms have more resources and stronger incentives to lobby the government for special treatment (e.g., the firm's city to be connected by HSR lines) (Zingales, 2017).

TABLE 8 Regression results from the instrumental variable (IV) approach (for the second stage).

	(1)	(2)	(3)
	ETR	*CETR*	*TAX_factor*
HSR_{pred}	0.049***	0.055**	0.345**
	(3.127)	(2.385)	(2.230)
SIZE	−0.004	−0.005	−0.006
	(−1.502)	(−1.282)	(−0.213)
LEV	−0.063***	0.063***	0.945***
	(−7.003)	(3.993)	(7.227)
ROE	0.016	−0.324***	−4.423***
	(1.603)	(−10.878)	(−11.137)
Q	−0.004***	−0.006***	−0.060***
	(−4.190)	(−3.717)	(−3.980)
PPE	−0.048***	0.015	0.119
	(−4.531)	(0.874)	(0.943)
INTANG	−0.074**	0.005	0.091
	(−2.265)	(0.102)	(0.251)
INVENTORY	0.042**	0.058**	0.452**
	(2.438)	(2.256)	(2.458)
LOSS	−0.016***	−0.008	−0.148***
	(−4.007)	(−1.110)	(−3.037)
SOE	0.004	−0.003	0.037
	(0.599)	(−0.256)	(0.592)
TREND	0.003*	0.005**	0.037**
	(1.950)	(1.981)	(2.201)
Firm and year FE	Yes	Yes	Yes
Industry × T	Yes	Yes	Yes
Province × T	Yes	Yes	Yes
N	18,073	16,370	14,395
Adj R^2	0.003	0.038	0.204
F	16.49	18.29	41.15
First stage F value	12.379	10.102	10.272
p value of overidentification tests	0.882	0.939	0.534

Notes: This table presents the second stage results of the two-stage least squares regressions. The indicator $RAILWAY_{c,\,1962}$ denotes whether city *c* possessed a railway station in 1962. It is assigned a value of 1 if the city had such a station and 0 otherwise. The interaction of $RAILWAY_{c,\,1962}$ and year dummies serves as the instrumental variable for the initiation of HSR services. HSR_{pred} is the predicted value of HSR derived from the first-stage regression. In the first stage, we regress the initiation of HSR services on the interaction between $RAILWAY_{c,\,1962}$ and *Year dummies*, along with other control variables. In the second stage, we regress our tax avoidance measures on HSR_{pred}, alongside other controls. All other variables are defined in Appendix A. The *t*-statistics based on robust standard errors clustered at the firm level are in parentheses.

*, ** and *** indicate statistical significance for a two-tailed test at the 10%, 5% and 1% levels, respectively.

TABLE 9 HSR and tax avoidance: the effect of insiders' rent-extraction propensity.

	ETR			*CETR*			*TAX_factor*		
	(1)	(2)	*DIFF*	(3)	(4)	*DIFF*	(3)	(4)	
Proxies for insiders' rent-extraction propensity	***High propensity***	***Low propensity***		***High propensity***	***Low propensity***		***High propensity***	***Low propensity***	***DIFF***
RPT	0.016***	0.005	0.099*	0.021**	−0.002	0.000**	0.146**	−0.002	0.000***
	(2.755)	(0.761)		(2.143)	(−0.160)		(2.517)	(−0.020)	
ORECTA	0.015***	0.009	0.040**	0.020**	−0.001	0.000**	0.176***	−0.011	0.000**
	(2.935)	(1.541)		(2.240)	(−0.066)		(3.149)	(−0.169)	
DIVERGENCE	0.010*	0.006	0.197	0.021***	−0.000	0.000**	0.160***	−0.037	0.000**
	(1.955)	(1.430)		(2.753)	(−0.004)		(3.138)	(−0.711)	
HHI10	0.012***	0.005	0.070*	0.014**	0.002	0.098*	0.121***	0.009	0.048**
	(2.937)	(1.094)		(2.046)	(0.260)		(2.787)	(0.148)	
BUSYDIRECT	0.013**	0.005	0.060*	0.014*	0.004	0.140	0.136**	0.013	0.039**
	(2.448)	(1.092)		(1.665)	(0.461)		(2.521)	(0.253)	
CEO_OWN	0.036**	0.005	0.000**	0.094**	0.004	0.000**	0.034	0.045	0.461
	(2.137)	(1.432)		(2.502)	(0.558)		(0.209)	(1.190)	
BIG10	0.008	0.007	0.486	0.013*	0.002	0.000**	0.093*	0.006	0.073*
	(1.595)	(1.631)		(1.698)	(0.281)		(1.887)	(0.124)	

Note: This table presents the results of OLS regressions that examine the effect of HSR line openings on corporate tax avoidance for different sub-samples. The full sample is divided into different sub-samples based on the median values of the proxies for rent-extraction propensity. *RPT* and *ORECTA* are the ratio of the abnormal related party transactions and abnormal other receivables to the total assets. *DIVERGENCE* is the difference between control right and cash flow right of controlling shareholder. *HHI10* is Herfindahl index of shareholdings of the 10 largest shareholders. *BUSYDIRECT* and *CEO_OWN* are the average number of directorships held by each independent director and the equity stake that the CEO holds. *BIG10* is a dummy variable that equals one if the firm is audited by Top 10 auditors based on an audit firm's Chinese revenues and zero otherwise. For brevity, only the coefficients of HSR are reported. The control variables are the same as specified in Table 4. All models include firm and year fixed effect, industry-by-year trend and province-by-province trend. *DIFF* denotes the *p* value for testing the equality of coefficients on *HSR* across groups. We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.

***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

difference in coefficients on *HSR* between the two groups is statistically significant. The tests using other proxies for rent-extraction propensity and tax avoidance provide similar results consistent with H2.

Overall, this cross-sectional evidence lends support to the narrative that the reduced information acquisition costs after the initiation of HSR services dissuade insiders, especially those from firms suffering severe agency problems, from pursuing rent-extraction-induced corporate tax avoidance.

4.4.2 | Additional tests

In extending the analysis to explore the underlying mechanism through which the launch of HSR services affects firms' tax avoidance, we follow prior research by undertaking a mediation analysis (Baron & Kenny, 1986; Gimbar et al., 2016; Wheeler, 2019). Extensive prior research implies that income shifting is the primary technique through which insiders in China siphon private benefits and avoid taxes (Bauer et al., 2020; Lin et al., 2018). If the lower tax avoidance that ensues after the launch of HSR services stems from insiders having weaker incentives to extract rents under the guise of tax avoidance, we should observe a lower level of income shifting after the launch of HSR services and the relation between the launch of HSR services and tax avoidance should be attenuated after controlling for income shifting. To test these conjectures, we follow Lin et al. (2018) by gaging income-shifting opportunities with the product of the difference between the highest and lowest statutory tax rates faced by any group member (i.e., the rate range) and the aggregate amount of related transactions (scaled by lagged assets) (*Shifting*). Afterward, we examine the links between the launch of HSR services, income shifting and tax avoidance. Table 10 reports the results. As shown in Columns (2), (4) and (6), the coefficients on *HSR* become smaller when we control for *Shifting*. Moreover, in Column (7), we find evidence suggesting that income shifting subsides after the opening of HSR services. Collectively, these results imply that income shifting plays a mediating role on the relation between the launch of HSR services and tax avoidance, providing some empirical support for the agency perspective explanation for our findings.

To further probe the role that corporate site visits play in mitigating rent-extraction-motivated tax avoidance, we examine the questions posed by external stakeholders during their visits. We posit that the convenience of site visits afforded by the introduction of HSR services, and consequently, the stricter oversight, matters more for firms whose external stakeholders are more concerned about insiders' motives for engaging in tax avoidance. In China, listed companies are obligated to promptly report all queries received during external stakeholders' site visits, along with their responses. The presence of the term "tax" in these reports suggests that stakeholders pay close attention to the company's tax strategies. Given that income shifting via *RPT*s is a common tactic for rent extraction and tax avoidance, we also search for the term "related party transaction" when classifying firms. We treat those with disclosed questions involving "tax" or "related party transaction" as having a higher perceived risk of exploiting tax strategies to facilitate rent extraction.

For this analysis, we reexamine the impact of the introduction of HSR lines on corporate tax avoidance, focusing on a group of firms that underwent site visits by external stakeholders in the preceding year. We categorize these firms into two groups depending on whether their disclosed queries mention "tax" or "related party transaction." Firms with such queries are labeled "Concern = 1," implying heightened concern over their tax strategies, whereas those without such queries are placed in the "Concern = 0" group. In Table 11, we report the regression results for both groups. Consistent with expectations, we find that the "Concern = 1" group has significantly higher HSR coefficients than the "Concern = 0" group. Moreover, these differences are statistically significant when we measure tax avoidance with *ETR* or *Tax_factor*. This evidence implies that the introduction of HSR enhances stakeholder monitoring, particularly for firms previously flagged for potential abuses in tax strategy.

TABLE 10 Additional test for the effect of agency cost: the mediation analysis.

	ETR		*CETR*		*TAX_factor*		*Shifting*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HSR	0.011***	0.007*	0.010*	0.010	0.082**	0.073*	−0.003**
	(3.059)	(1.660)	(1.913)	(1.476)	(2.314)	(1.722)	(−2.100)
Shifting		−0.025		−0.127***		−1.481***	
		(−0.857)		(−2.768)		(−4.425)	
SIZE	−0.004*	−0.001	−0.006	0.000	−0.011	0.018	0.009***
	(−1.663)	(−0.260)	(−1.412)	(0.032)	(−0.371)	(0.445)	(5.247)
LEV	−0.062***	−0.061***	0.062***	0.098***	0.949***	1.085***	0.031***
	(−6.860)	(−4.659)	(3.921)	(4.298)	(7.215)	(6.338)	(5.755)
ROE	0.023**	0.038***	−0.319***	−0.334***	−4.361***	−4.266***	0.013**
	(2.351)	(2.925)	(−10.601)	(−7.425)	(−10.851)	(−7.249)	(2.246)
Q	−0.004***	−0.003**	−0.006***	−0.006**	−0.059***	−0.066**	−0.000
	(−4.084)	(−2.117)	(−3.843)	(−2.155)	(−4.029)	(−2.520)	(−0.722)
PPE	−0.046***	−0.060***	0.015	0.032	0.139	0.112	−0.006
	(−4.444)	(−4.252)	(0.920)	(1.274)	(1.097)	(0.620)	(−0.947)
INTANG	−0.079**	−0.107**	0.005	−0.014	0.112	−0.056	0.004
	(−2.439)	(−2.437)	(0.105)	(−0.220)	(0.305)	(−0.111)	(0.257)
INVENTORY	0.041**	0.035	0.057**	0.067**	0.443**	0.444*	−0.013
	(2.350)	(1.577)	(2.261)	(2.034)	(2.395)	(1.875)	(−1.308)
LOSS	−0.015***	−0.022***	−0.009	−0.017	−0.150***	−0.212***	−0.001
	(−3.901)	(−4.529)	(−1.208)	(−1.579)	(−3.055)	(−2.879)	(−0.704)
SOE	0.004	0.014	−0.002	0.009	0.032	−0.016	−0.007**
	(0.590)	(1.370)	(−0.216)	(0.507)	(0.515)	(−0.139)	(−2.536)
TREND	0.000	−0.002	0.001	0.001	0.013	0.005	−0.000
	(0.088)	(−1.187)	(0.605)	(0.379)	(1.300)	(0.354)	(−0.429)
Firm and year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	17,834	8760	16,173	7804	14,208	7409	9126
Adj R^2	0.405	0.446	0.250	0.247	0.460	0.431	0.374
F	12.40	7.049	11.99	12.64	25.32	21.13	6.623

Note: This table presents the results of OLS regressions that examine the relationship among the launch of HSR services, income shifting and corporate tax avoidance. *Shifting* is the tax-motivated income-shifting opportunities, which is the product of the difference between the highest and lowest statutory tax rates among group members (i.e., the rate range) and the aggregate amount of related party transactions (scaled by lagged assets of the group). All models include firm and year fixed effect, industry-by-year trend and province-by-year trend. We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.

***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

TABLE 11 Additional test for the effect of agency cost: the evidence from site visits.

	ETR		*CETR*		*TAX_factor*	
	(1)	(2)	(3)	(4)	(5)	(6)
	Concern = 1	Concern = 0	Concern = 1	Concern = 0	Concern = 1	Concern = 0
HSR	0.050**	0.004	0.033	0.012	0.137	−0.074
	(2.377)	(0.479)	(0.949)	(0.577)	(0.900)	(−0.726)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm and year FE	Yes	Yes	Yes	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes	Yes	Yes	Yes
N	1394	3315	1317	3035	1268	2924
Adj R^2	0.556	0.403	0.383	0.264	0.667	0.599
F	41.25	4.699	14.84	5.505	29.86	34.38
p value of differences in HSR	0.000***		0.190		0.020**	

Notes: This table re-examines the effect of HSR line openings on corporate tax avoidance, concentrating on a subset of firms that underwent site visits by external stakeholders in the previous year. These firms are split into two groups depending on whether their disclosed queries during the visits involved terms like "tax" or "related party transaction." Firms with such queries are categorized as "Concern = 1," whereas those without them fall into the "Concern = 0" group. All models include firm and year fixed effect, industry-by-year trend and province-by-year trend. We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.
***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

4.5 | Alternative explanations

In this section, we examine several alternative explanations for the finding that firms' tax avoidance declines after the opening of HSR lines. For starters, firms may experience stricter monitoring from tax authorities afterward, leading to lower tax avoidance. Convenient transportation can result in tax authorities imposing tougher enforcement (Kubick et al., 2017), which may, in turn, deter firms' tax avoidance (Hoopes et al., 2012; Kubick et al., 2017). However, for several reasons, we expect that the launch of HSR services will have only a minimal impact on tax enforcement severity. First, given that the monitoring activities are ordinarily handled by city-level tax offices in China (Cai & Liu, 2009), the improvement in inter-city transportation infrastructure should be almost irrelevant to tax authorities' monitoring activities. Second, because local taxation authorities in China primarily focus on ensuring that they meet their annual tax collection target, instead of maximizing current tax revenues, the opening of HSR lines should hardly matter to corporate tax enforcement (Deng & Luo, 2011). Nonetheless, we consider the empirical relevance of this competing explanation by comparing the impact of HSR services becoming available on tax avoidance in regions with varying levels of tax enforcement. If the reduction in tax avoidance reflects stricter monitoring by tax authorities, then this effect should be magnified in regions with weak tax enforcement before the launch of HSR services. We follow Xu et al. (2011) by measuring province-level tax enforcement with the ratio of actual tax payments over the estimated tax payments. In untabulated analysis, we find that there is no significant shift in province-level tax enforcement after the opening of HSR lines and there is no perceptible difference in the role that the introduction of HSR services plays in corporate tax avoidance across provinces with different levels of original tax enforcement. Accordingly, it would be hard to accept that shifts in tax enforcement are responsible for our core evidence.

Next, we address that it is plausible that firms pay more taxes after the launch of HSR services, given the local government's higher demand for tax revenue to pay for their share of the construction costs. Given that local firms

directly benefit from the launch of HSR services, local governments may insist that they subsidize the construction costs. To test this alternative explanation, we analyze whether the tax impact of HSR is stronger when local governments have large fiscal deficits, which engender higher demand for tax revenue. However, we do not find any evidence of a stronger effect of the opening of HSR lines on tax avoidance in cities with large fiscal deficits, implying that our results are not driven by government incentives.

Additionally, we examine whether improving transportation infrastructure alleviates financial constraints, reducing firms' incentive to conserve cash through aggressive tax planning. Prior work implies that equity financing is cheaper when the distance between investors and firms is short because geographic proximity lowers information asymmetry (Butler, 2008; Degryse & Ongena, 2005). Moreover, recent tax research suggests that financially constrained firms are eager to increase cash tax savings both because it is a cheaper way to raise cash compared with traditional financing sources, which has little impact on their operations (Edwards et al., 2016; Law & Mills, 2015). Accordingly, the opening of HSR lines is also likely to benefit firms by lowering their cost of capital and relaxing their financial constraints, which, in turn, lessens demand for tax avoidance. To analyze this conjecture, we split our full sample into paired sub-groups based on the year-median values of the proxies for the level of financial constraints and rerun the main regressions. This alternative explanation predicts that the negative impact of the launch of HSR services on tax avoidance intensifies for firms experiencing financial constraints beforehand.

We follow Hadlock and Pierce (2010) by gaging financial constraints with a financial constraint index (*SA index*).[19] In untabaluted analysis, we find that the coefficients on *HSR* are only positive and statistically significant for firms enjoying a low level of financial constraints. These results do not reconcile with the relaxation of financial constraints explanation; rather, they imply that insiders' rent-extraction incentives are stronger in firms with a low level of financial constraints because these firms have more resources available for diversion.[20]

4.6 | HSR, tax avoidance and future performance

In the previous sections, we report evidence implying that firms engage in less tax avoidance after the opening of HSR lines because they no longer rely heavily on complex tax planning to provide cover for their diversionary activities. In our final analysis, we compare firms' values before and after the launch of HSR services. If the reduction in tax avoidance represents lower rent extraction, we should observe that firm value rises after the launch of HSR services. To evaluate this conjecture, we estimate the following regression model:

$$Perf = \beta_0 + \beta_1 Pre\_TAX_{it} + \gamma CONTROLS_{i,t} + MACRO_{c,2003} \times \omega_t + \nu_i + \omega_t + \delta_p \times \omega_t + \eta_k \times \omega_t + \varepsilon_{it} \quad (6)$$

where *Perf* is proxied by Tobin's *Q* and *ROE*. *Pre_TAX* is the level of tax avoidance mitigated by the opening of HSR lines, which we compute as follows (Core et al., 1999; Mao, 2021):

$$Pre\_TAX_{i,t} = \hat{\beta}_1 HSR_{i,t} \quad (7)$$

where the estimated coefficients on the *HSR* variables ($\hat{\beta}_1$) are those reported in Table 4. We predict that firm value *Q* in year *t* and *ROE* in year $t + 1$ are positively associated with *Pre_TAX* in year *t*.

As shown in Table 12, the coefficients on *Pre_TAX* enter positively across all columns. The results suggest that a one-standard deviation increase in *Pre_TAX* (the standard deviations are approximately 0.434%, 0.439% and 3.504%,

[19] We also follow Almeida et al. (2004) by specifying firm size, age and cash dividends as proxies for financial constraints and find similar results.

[20] In another unreported test, we also examine whether the fall in aggressive tax planning stems firms the increased investment of firms in cities with high tax rates. To dispel this alternative explanation, we exclude firms that increase their investment or sales outside their home cities after the opening of HSR lines and find that our main results remain the same.

TABLE 12 HSR, tax avoidance and future performance.

	Q_{it}	Q_{it}	Q_{it}	ROE_{it+1}	ROE_{it+1}	ROE_{it+1}
	(1)	(2)	(3)	(4)	(5)	(6)
Pre_ETR	4.750*			0.809***		
	(1.807)			(2.757)		
Pre_CETR		6.236**			0.901***	
		(2.262)			(3.123)	
Pre_TAX_factor			0.797**			0.134***
			(2.273)			(3.439)
SIZE	−0.317***	−0.306***	−0.290***	−0.019***	−0.021***	−0.020***
	(−13.615)	(−12.932)	(−11.686)	(−7.545)	(−8.231)	(−7.218)
LEV	−0.027	−0.050	−0.087	0.094***	0.094***	0.088***
	(−0.316)	(−0.590)	(−0.977)	(8.193)	(7.710)	(6.665)
ROE	1.995***	2.094***	2.042***	0.310***	0.316***	0.327***
	(11.859)	(12.788)	(10.799)	(10.235)	(9.712)	(8.594)
Q_{it}				0.006***	0.006***	0.007***
				(5.347)	(5.236)	(4.988)
Q_{it-1}	0.409***	0.418***	0.416***			
	(31.305)	(32.014)	(30.831)			
PPE	0.410***	0.399***	0.402***	−0.035***	−0.033***	−0.037***
	(3.951)	(3.776)	(3.728)	(−3.451)	(−3.160)	(−3.236)
INTANG	0.882***	0.820***	0.848***	−0.063*	−0.086**	−0.101***
	(3.522)	(3.379)	(3.310)	(−1.777)	(−2.463)	(−3.029)
INVENTORY	0.181	0.166	0.251*	−0.006	0.004	0.007
	(1.381)	(1.262)	(1.835)	(−0.427)	(0.288)	(0.466)
LOSS	−0.085	−0.165	−0.093	0.004	0.003	0.005
	(−0.821)	(−1.205)	(−0.710)	(0.857)	(0.656)	(0.932)
SOE	−0.074	−0.092*	−0.085	−0.010*	−0.009*	−0.010**
	(−1.504)	(−1.783)	(−1.585)	(−1.797)	(−1.741)	(−2.116)
BIND	0.002	0.001	0.001	−0.000	−0.000	−0.000
	(0.901)	(0.320)	(0.375)	(−1.178)	(−1.193)	(−0.126)
DUAL	−0.025	−0.010	−0.008	−0.006**	−0.003	−0.004
	(−1.056)	(−0.408)	(−0.329)	(−2.254)	(−1.211)	(−1.368)
TOP1	−0.005***	−0.005***	−0.004***	0.000	0.000	0.000
	(−3.744)	(−3.679)	(−3.183)	(1.068)	(0.416)	(0.488)
TREND	0.003	0.009	0.007	−0.000	0.000	0.000
	(0.414)	(1.083)	(0.831)	(−0.206)	(0.109)	(0.448)
Firm and year FE	Yes	Yes	Yes	Yes	Yes	Yes
$MACRO_{2003} \times T$	Yes	Yes	Yes	Yes	Yes	Yes
Industry × *T*	Yes	Yes	Yes	Yes	Yes	Yes
Province × *T*	Yes	Yes	Yes	Yes	Yes	Yes

(Continues)

TABLE 12 (Continued)

	Q_{it} (1)	Q_{it} (2)	Q_{it} (3)	ROE_{it+1} (4)	ROE_{it+1} (5)	ROE_{it+1} (6)
N	14,107	13,132	12,626	14,472	13,397	11,703
Adj R^2	0.701	0.707	0.706	0.477	0.493	0.501
F	103.7	104.2	92.28	32.30	28.21	27.99

Note: This table presents the results of OLS regressions that examine the effect of mitigated tax avoidance due to the opening of HSR lines on firm value. *Q* is the sum of market value of tradable shares, book value of non-tradable shares and liabilities scaled by book value of total assets. *Pre_ETR*, *Pre_CETR and Pre_TAX_factor*, which are proxies for the level of tax avoidance mitigated by the opening of HSR lines, are estimated with Equation (7). All models include firm and year fixed effect, industry-by-year trend and province-by-year trend. We report *t* statistics in parentheses under the coefficients. Standard errors are clustered by firm. Definitions of all variables are presented in Appendix A.
***, ** and * denote significance at the 1%, 5% and 10% levels, respectively.

respectively) is associated with 2.06%, 2.737% and 2.79% increases in Tobin's *Q* (the coefficients for *Pre_TAX* are 4.750, 6.236 and 0.797 in Columns 1, 2 and 3, respectively) and 0.35%, 0.40% and 0.47% growth in ROE (the coefficients for ROE are 0.809, 0.901 and 0.134 in Columns 4, 5 and 6, respectively). Altogether, the results in this section suggest that the reduced tax avoidance due to the opening of HSR lines is value-enhancing, which provides further evidence reinforcing our main hypotheses.

5 | CONCLUSION

We exploit the launch of HSR services in China as a unique setting to examine how geographic proximity stemming from transportation infrastructure development shapes firms' tax avoidance. We find that the opening of HSR lines facilitates corporate site visits by external stakeholders and leads to a significant drop in firms' tax avoidance. These results hold for a variety of alternative measures and robustness checks. Moreover, we also report that the impact of HSR line openings on tax avoidance intensifies for firms in which insiders have high rent-extraction propensity. Additional analyses reveal that firms undertake less tax avoidance via income shifting and the reduced tax avoidance is due to the increase in the intensity of site visits that are undertaken by institutional investors and financial analysts. Finally, we show that firm value rises after HSR services become available. Collectively, our results suggest that reducing costs associated with information acquisition, particularly the acquisition of soft information, narrows the scope for diversionary activities. Accordingly, insiders become less eager to pursue complex tax planning to provide cover for their private benefits extraction.

Leveraging the launch of HSR services as a natural experiment, our staggered DID design helps dispel endogeneity threats to reliable inference to improve identification on whether insiders implement complex tax strategies to facilitate the diversion of corporate resources. We also document a major positive externality stemming from the development of transportation infrastructure: an increase in government tax revenue. As such, our findings may have important practical implications for policymakers.

ACKNOWLEDGMENTS

We thank Chen Chen, Simon Fung, Ferdinand Gul, Bin Ke, Sunyoung Kim, Kenny Lin, Arthur Stenzel, Nancy Su, Zitian Wang, Anne Wyatt, Zhifeng Yang, Yue Zhang and participants in the 2021 EAA Virtual Annual Congress, as well as workshop participants at Deakin University, Monash University and Lingnan University, for providing valuable insights on an earlier version of this paper. We also gratefully acknowledges the financial support from National Natural

Science Foundation of China (grant number 72162019, 72102010), Social Science Fund of Jiangxi Province (grant number 21YJ30). All authors are listed in alphabetic order by the surname and have contributed equally.

DATA AVAILABILITY STATEMENT

Information Acquisition and Tax Avoidance: Evidence from a Natural Experiment Lyu Fan, Caiyue Ouyang, Jeffrey Pittman, Jiacai Xiong and Jun Yao. (1) A description of which author(s) handled the data and conducted the analyses. All authors were extensively involved in the research design, including the approach to developing the sample and collecting the data. Jiacai Xiong and Caiyue Ouyang were specifically involved in handling the data and conducting the data analyses. (2) A detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained and the instrument used to generate the data (e.g., for surveys or experiments). We began our data collection on March 4, 2019 by downloading the financial data from the China Stock Market and Accounting Research (CSMAR) database and retrieve the variables necessary for computing the statutory tax rate from the WIND database. Data on corporate site visits are from the Chinese Research Data Services (CNRDS) platform. We manually collect information about the opening of HSR lines from the website of China's National Railway Administration (http://www.12306.cn). The sample period begins in 2005, which is three years before 2008 when the first HSR line was opened. Our initial sample includes all firms listed on the Shanghai Stock Exchange or the Shenzhen Stock Exchange, amounting to 52,500 firm-year observations from 2005 to 2022. Next, we exclude: (i) 364 firm-years from the financial industry; (ii) 5,970 observations with effective tax rates larger than 1 or less than 0; (iii) 2,889 observations with negative pretax income; (iv) 20,548 firm-years from the four municipalities (i.e. Beijing, Tianjin, Shanghai and Chongqing) and 28 provincial capital cities to confront any lingering endogeneity concerns; and (v) 4,895 firm-years with missing values for the control variables used in our analyses. Our final sample contains 17,834 firm-year observations representing 2,210 unique firms. (3) If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The data are not obtained from an organization on a proprietary basis. (4) The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. The programs are available upon requested. (5) An assurance that the data and programs will be maintained by at least one author for at least 6 years, consistent with National Science Foundation guidelines. The data and programs will be maintained by Jiacai Xiong, Caiyue Ouyang and Jun Yao for at least 6 years.

ORCID

Caiyue Ouyang https://orcid.org/0000-0002-7428-2886
Jiacai Xiong https://orcid.org/0000-0003-2786-6448

REFERENCES

Allen, A., Francis, B. B., Wu, Q., & Zhao, Y. (2016). Analyst coverage and corporate tax aggressiveness. *Journal of Banking and Finance*, *73*, 84–98.

Almeida, H., Campello, M., & Weisbach, M. S. (2004). The cash flow sensitivity of cash. *Journal of Finance*, *59*(4), 1777–1804.

Armstrong, C. S., Blouin, J. L., Jagolinzer, A. D., & Larcker, D. F. (2015). Corporate governance, incentives, and tax avoidance. *Journal of Accounting and Economics*, *60*(1), 1–17.

Armstrong, C. S., Blouin, J. L., & Larcker, D. F. (2012). The incentives for tax planning. *Journal of Accounting and Economics*, *53*(1–2), 391–411.

Atwood, T. J., & Lewellen, C. (2019). The complementarity between tax avoidance and manager diversion: Evidence from tax haven firms. *Contemporary Accounting Research*, *36*(1), 259–294.

Ayers, B. C., Ramalingegowda, S., & Yeung, P. E. (2011). Hometown advantage: The effects of monitoring institution location on financial reporting discretion. *Journal of Accounting and Economics*, *52*(1), 41–61.

Badertscher, B. A., Katz, S. P., & Rego, S. O. (2013). The separation of ownership and control and corporate tax avoidance. *Journal of Accounting and Economics*, *56*(2–3), 228–250.

Bae, K. H., Kang, J. K., & Kim, J. M. (2002). Tunneling or value added? Evidence from mergers by Korean business groups. *Journal of Finance*, *57*(6), 2695–2740.
Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395.
Bai, J., Fairhurst, D., & Serfling, M. (2020). Employment protection, investment, and firm growth. *Review of Financial Studies*, *33*(2), 644–688.
Baik, B., Kang, J. K., & Kim, J. M. (2010). Local institutional investors, information asymmetries, and equity returns. *Journal of Financial Economics*, *97*(1), 81–106.
Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.
Bauer, A. M., Fang, J., Pittman, J., Zhang, Y., & Zhao, Y. (2020). How aggressive tax planning facilitates the diversion of corporate resources: Evidence from path analysis. *Contemporary Accounting Research*, *37*(3), 1882–1913.
Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics*, *99*(3), 435–448.
Bernard, A. B., Moxnes, A., & Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy*, *127*(2), 639–688.
Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The impact of venture capital monitoring. *Journal of Finance*, *71*(4), 1591–1622.
Blaylock, B. S. (2016). Is tax avoidance associated with economically significant rent extraction among U.S. firms? *Contemporary Accounting Research*, *33*(3), 1013–1043.
Bradshaw, M., Liao, G., & Ma, M. S. (2019). Agency costs and tax planning when the government is a major shareholder. *Journal of Accounting and Economics*, *67*(2–3), 255–277.
Butler, A. W. (2008). Distance still matters: Evidence from municipal bond underwriting. *Review of Financial Studies*, *21*(2), 763–784.
Cai, H., & Liu, Q. (2009). Competition and corporate tax avoidance: Evidence from Chinese industrial firms. *The Economic Journal*, *119*(537), 764–795.
Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, *134*(3), 1405–1454.
Cao, J., Wang, H., & Zhou, S. (2022). Soft activism and corporate dividend policy: Evidence from institutional investors site visits. *Journal of Corporate Finance*, *75*, 102221.
Chan, K. H., Lin, K. Z., & Mo, P. L. (2010). Will a departure from tax-based accounting encourage tax noncompliance? Archival evidence from a transition economy. *Journal of Accounting and Economics*, *50*(1), 58–73.
Chan, K. H., Mo, P. L. L., & Tang, T. (2016). Tax avoidance and tunneling: Empirical analysis from an agency perspective. *Journal of International Accounting Research*, *15*(3), 49–66.
Chemmanur, T. J., Hull, T. J., & Krishnan, K. (2016). Do local and international venture capitalists play well together? The complementarity of local and international venture capitalists. *Journal of Business Venturing*, *31*(5), 573–594.
Chen, C. X., Lu, H., & Sougiannis, T. (2012). The agency problem, corporate governance, and the asymmetrical behavior of selling, general, and administrative costs. *Contemporary Accounting Research*, *29*(1), 252–282.
Chen, D., Ma, Y., Martin, X., & Michaely, R. (2022). On the fast track: Information acquisition costs and information production. *Journal of Financial Economics*, *143*(2), 794–823.
Chen, H., Qu, Y., Shen, T., Wang, Q., & Xu, D. X. (2022). The geography of information acquisition. *Journal of Financial and Quantitative Analysis*, *57*(6), 2251–2285.
Chen, K., Chi, J. D., & Guo, W. (2021). *Information super-railway? Transportation infrastructure and information production in financial markets* [Working paper].
Chen, R., El Ghoul, S., Guedhami, O., Wang, H., & Yang, Y. (2022). Corporate governance and tax avoidance: Evidence from US cross-listing. *The Accounting Review*, *97*(7), 49–78.
Chen, H. W., Tang, S., Wu, D. H., & Yang, D. G. (2021). The political dynamics of corporate tax avoidance: The Chinese experience. *The Accounting Review*, *96*(5), 157–180.
Chen, K.-P., & Chu, C. C. (2005). Internal control versus external manipulation: A model of corporate income tax evasion. *Rand Journal of Economics*, *36*, 151–164.
Chen, N. X., Chiu, P. C., & Shevlin, T. (2018). Do analysts matter for corporate tax planning? Evidence from a natural experiment. *Contemporary Accounting Research*, *35*(2), 794–829.
Chen, S., Chen, X., Cheng, Q., & Shevlin, T. (2010). Are family firms more tax aggressive than non-family firms? *Journal of Financial Economics*, *95*(1), 41–61.
Chen, S., Huang, Y., Li, N., & Shevlin, T. (2019). How does quasi-indexer ownership affect corporate tax planning? *Journal of Accounting and Economics*, *67*(2–3), 278–296.

Chen, T., & Lin, C. (2017). Does information asymmetry affect corporate tax aggressiveness? *Journal of Financial and Quantitative Analysis*, *52*(5), 2053–2081.
Cheng, C. S. A., Huang, H. H., Li, Y., & Stanfield, J. (2012). The effect of hedge fund activism on corporate tax avoidance. *The Accounting Review*, *87*(5), 1493–1526.
Cheng, Q., Du, F., Wang, B. Y., & Wang, X. (2019). Do corporate site visits impact stock prices? *Contemporary Accounting Research*, *36*(1), 359–388.
Cheng, Q., Du, F., Wang, X., & Wang, Y. (2016). Seeing is believing: Analysts' corporate site visits. *Review of Accounting Studies*, *21*(4), 1245–1286.
Cheung, Y.-L., Rau, P. R., & Stouraitis, A. (2006). Tunneling, propping, and expropriation: Evidence from connected party transactions in Hong Kong. *Journal of Financial Economics*, *82*(2), 343–386.
Chung, S. G., Goh, B. W., Lee, J., & Shevlin, T. (2019). Corporate tax aggressiveness and insider trading. *Contemporary Accounting Research*, *36*(1), 230–258.
Claessens, S., Djankov, S., Fan, J. P., & Lang, L. H. (2002). Disentangling the incentive and entrenchment effects of large shareholdings. *Journal of Finance*, *57*(6), 2741–2771.
Core, J. E., Holthausen, R. W., & Larcker, D. F. (1999). Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics*, *51*(3), 371–406.
Cornaggia, J., Mao, Y., Tian, X., & Wolfe, B. (2015). Does banking competition affect innovation? *Journal of Financial Economics*, *115*(1), 189–209.
Cui, C., & Li, L. S.-Z. (2019). High-speed rail and inventory reduction: Firm-level evidence from China. *Applied Economics*, *51*(25), 2715–2730.
Degryse, H., & Ongena, S. (2005). Distance, lending relationships, and competition. *Journal of Finance*, *60*(1), 231–266.
Deng, Y. H., & Luo, T. (2011). Tax revenue manipulation by local taxation administration in China. *Asia-Pacific Journal of Accounting and Economics*, *18*(1), 61–76.
Desai, M. A. (2005). The degradation of reported corporate profits. *Journal of Economic Perspectives*, *19*(4), 171–192.
Desai, M. A., & Dharmapala, D. (2006). Corporate tax avoidance and high-powered incentives. *Journal of Financial Economics*, *79*(1), 145–179.
Desai, M. A., & Dharmapala, D. (2009a). Corporate tax avoidance and firm value. *Review of Economics and Statistics*, *91*(3), 537–546.
Desai, M. A., & Dharmapala, D. (2009b). Earnings management, corporate tax shelters, and book-tax alignment. *National Tax Journal*, *62*, 169–186.
Desai, M. A., Dyck, A., & Zingales, L. (2007). Theft and taxes. *Journal of Financial Economics*, *84*(3), 591–623.
Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *Quarterly Journal of Economics*, *131*(2), 799–858.
Dong, X., Zheng, S., & Kahn, M. E. (2020). The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, *115*, 103212.
Duflo, E., & Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, *122*(2), 601–646.
Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, *79*(4), 1407–1440.
Dyreng, S. D., Hanlon, M., & Maydew, E. L. (2008). Long-run corporate tax avoidance. *The Accounting Review*, *83*(1), 61–82.
Dyreng, S. D., & Lindsey, B. P. (2009). Using financial accounting data to examine the effect of foreign operations located in tax havens and other countries on U.S. multinational firms' tax rates. *Journal of Accounting Research*, *47*(5), 1283–1316.
Dyreng, S. D., & Maydew, E. L. (2018). Virtual issue on tax research. *Journal of Accounting Research*, *56*(1), 3–4.
Edwards, A., Schwab, C., & Shevlin, T. (2016). Financial constraints and cash tax savings. *The Accounting Review*, *91*(3), 859–881.
Ellis, J., 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.
El Ghoul, S., Guedhami, O., Ni, Y., Pittman, J., & Saadi, S. (2013). Does information asymmetry matter to equity pricing? Evidence from firms' geographic location. *Contemporary Accounting Research*, *30*(1), 140–181.
Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's national trunk highway system. *Review of Economic Studies*, *81*(3), 1046–1070.
Fich, E. M., & Shivdasani, A. (2005). Are busy boards effective monitors? *Journal of Finance*, *61*, 689–724.
Frank, M. M., Lynch, L. J., & Rego, S. O. (2009). Tax reporting aggressiveness and its relation to aggressive financial reporting. *The Accouning Review*, *84*(2), 467–496.
Gallemore, J., Maydew, E. L., & Thornock, J. R. (2014). The reputational costs of tax avoidance. *Contemporary Accounting Research*, *31*(4), 1103–1133.
Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics*, *121*(3), 931–972.
Gimbar, C., Hansen, B., & Ozlanski, M. E. (2016). The effects of critical audit matter paragraphs and accounting standard precision on auditor liability. *The Accounting Review*, *91*(6), 1629–1646.

Giroud, X. (2013). Proximity and investment: Evidence from plant-level data. *The Quarterly Journal of Economics*, *128*(2), 861–915.
Givoni, M. (2006). Development and impact of the modern high-speed train: A review. *Transport Reviews*, *26*(5), 593–611.
Glaeser, E. L. (1999). Learning in cities. *Journal of Urban Economics*, *46*(2), 254–277.
Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.
Gormley, T. A., & Matsa, D. A. (2014). Common errors: How to (and not to) control for unobserved heterogeneity. *Review of Financial Studies*, *27*(2), 617–661.
Gul, F. A., Kim, J.-B., & Qiu, A. A. (2010). Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China. *Journal of Financial Economics*, *95*(3), 425–442.
Hadlock, C. J., & Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the KZ index. *Review of Financial Studies*, *23*(5), 1909–1940.
Han, B., Kong, D., & Liu, S. (2018). Do analysts gain an informational advantage by visiting listed companies? *Contemporary Accounting Research*, *35*(4), 1843–1867.
He, X., Zhi, J., Pittman, J., & Rui, O. M. (2019). *The impact of transportation infrastructure development on auditor effort and audit quality: Evidence from a natural experiment* (SSRN 3405107). SSRN.
Henry, E., & Sansing, R. (2018). Corporate tax avoidance: Data truncation and loss firms. *Review of Accounting Studies*, *23*(3), 1042–1070.
Hoopes, J. L., Mescall, D., & Pittman, J. (2012). Do IRS audits deter corporate tax avoidance? *The Accounting Review*, *87*(5), 1603–1639.
Hou, Q., Liu, Q., Yang, Z., & Zhang, L. (2018). *China railway high-speed: The impact of transportation infrastructure on corporate innovation* (SSRN 3153965). SSRN.
Hutton, A. P., Marcus, A. J., & Tehranian, H. (2009). Opaque financial reports, R2, and crash risk. *Journal of Financial Economics*, *94*(1), 67–86.
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*(1), 267–306.
Jian, M., & Wong, T. J. (2010). Propping through related party transactions. *Review of Accounting Studies*, *15*(1), 70–105.
Jiang, G., Lee, C. M., & Yue, H. (2010). Tunneling through intercorporate loans: The China experience. *Journal of Financial Economics*, *98*(1), 1–20.
Jiang, X., & Yuan, Q. (2018). Institutional investors' corporate site visits and corporate innovation. *Journal of Corporate Finance*, *48*, 148–168.
Khan, M., Srinivasan, S., & Tan, L. (2017). Institutional ownership and corporate tax avoidance: New evidence. *The Accounting Review*, *92*(2), 101–122.
Khan, S., Park, S. J., Veliotis, S., & Wald, J. K. (2023). Director and officer liability and corporate tax avoidance. *Journal of Business Finance & Accounting*, *50*(7–8), 1338–1371.
Khurana, I. K., & Moser, W. J. (2013). Institutional shareholders' investment horizons and tax avoidance. *Journal of the American Taxation Association*, *35*(1), 111–134.
Kim, J.-B., Li, Y., & Zhang, L. (2011). Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics*, *100*(3), 639–662.
Knyazeva, A., Knyazeva, D., & Masulis, R. W. (2013). The supply of corporate directors and board independence. *Review of Financial Studies*, *26*(6), 1561–1605.
Kubick, T. R., Lockhart, G. B., Mills, L. F., & Robinson, J. R. (2017). IRS and corporate taxpayer effects of geographic proximity. *Journal of Accounting and Economics*, *63*(2–3), 428–453.
La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. *Journal of Finance*, *54*(2), 471–517.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (2000). Investor protection and corporate governance. *Journal of Financial Economics*, *58*(1–2), 3–27.
Lanis, R., & Richardson, G. (2011). The effect of board of director composition on corporate tax aggressiveness. *Journal of Accounting and Public Policy*, *30*(1), 50–70.
Law, K. K., & Mills, L. F. (2015). Taxes and financial constraints: Evidence from linguistic cues. *Journal of Accounting Research*, *53*(4), 777–819.
Lawrence, M., Bullock, R., & Liu, Z. (2019). *China's high-speed rail development*. World Bank Publications.
Lennox, C., Lisowsky, P., & Pittman, J. (2013). Tax aggressiveness and accounting fraud. *Journal of Accounting Research*, *51*(4), 739–778.
Lennox, C., Wu, X., & Zhang, T. (2016). The effect of audit adjustments on earnings quality: Evidence from China. *Journal of Accounting and Economics*, *61*(2–3), 545–562.

Li, O. Z., Liu, H., & Ni, C. (2017). Controlling shareholders' incentive and corporate tax avoidance: A natural experiment in China. *Journal of Business Finance and Accounting*, *44*(5–6), 697–727.

Li, P., Lu, Y., & Wang, J. (2020). The effects of fuel standards on air pollution: Evidence from China. *Journal of Development Economics*, *146*, 102488.

Lin, K. Z., Mills, L. F., Zhang, F., & Li, Y. (2018). Do political connections weaken tax enforcement effectiveness? *Contemporary Accounting Research*, *35*(4), 1941–1972.

Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from China's high speed railway system. *Journal of Urban Economics*, *98*, 98–123.

Lin, Y., Qin, Y., Sulaeman, J., Yan, J., & Zhang, J. (2023). Expanding footprints: The impact of passenger transportation on corporate locations. *Review of Finance*, *27*(3), 1119–1154.

Lin, Y., Xin, X., Zhang, L., & Zhang, Z. (2017). *Bank monitoring and corporate tax planning: Evidence from loan covenants* [Working paper]. National University of Singapore.

Liu, S., Dai, Y., & Kong, D. (2017). Does it pay to communicate with firms? Evidence from firm site visits of mutual funds. *Journal of Business Finance and Accounting*, *44*(5–6), 611–645.

Lo, A. W., Wong, R. M., & Firth, M. (2010). Tax, financial reporting, and tunneling incentives for income shifting: An empirical analysis of the transfer pricing behavior of Chinese-listed companies. *Journal of the American Taxation Association*, *32*(2), 1–26.

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*(2), 719–755.

Manzon, G. B., Jr., & Plesko, G. A. (2001). The relation between financial and tax reporting measures of income. *Tax Law Review*, *55*, 175.

Mao, Y. F. (2021). Managing innovation: The role of collateral. *Journal of Accounting & Economics*, *72*(1):101419.

Markle, K. (2016). A comparison of the tax-motivated income shifting of multinationals in territorial and worldwide countries. *Contemporary Accounting Research*, *33*(1), 7–43.

McGuire, S. T., Omer, T. C., & Wang, D. (2012). Tax avoidance: Does tax-specific industry expertise make a difference? *The Accounting Review*, *87*(3), 975–1003.

Möller, J., & Zierer, M. (2018). Autobahns and jobs: A regional study using historical instrumental variables. *Journal of Urban Economics*, *103*, 18–33.

O'Brien, P. C., & Tan, H. (2015). Geographic proximity and analyst coverage decisions: Evidence from IPOs. *Journal of Accounting and Economics*, *59*(1), 41–59.

Ouyang, C., Xiong, J., Liu, L., & Yao, J. (2024). Geographic proximity and trade credit: Evidence from a quasi-natural experiment. *Journal of Corporate Finance*, *84*, 102535.

Phillips, J. D. (2003). Corporate tax-planning effectiveness: The role of compensation-based incentives. *The Accounting Review*, *78*(3), 847–874.

Plesko, G. A. (2003). An evaluation of alternative measures of corporate tax rates. *Journal of Accounting and Economics*, *35*(2), 201–226.

Porcano, T. (1986). Corporate tax rates: Progressive, proportional, or regressive. *Journal of the American Taxation Association*, *7*(2), 17–31.

Redding, S. J., & Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. In *Handbook of* regional and urban economics (Vol. 5, pp. 1339–1398). Elsevier.

Seidman, J. K., & Stomberg, B. (2017). Equity compensation and tax avoidance: Disentangling managerial incentives from tax benefits and reexamining the effect of shareholder rights. *Journal of the American Taxation Association*, *39*(2), 21–41.

Shevlin, T. (1987). Taxes and off-balance-sheet financing: Research and development limited partnerships. *The Accounting Review*, *62*, 480–509.

Shevlin, T., Tang, T. Y., & Wilson, R. J. (2012). Domestic income shifting by Chinese listed firms. *Journal of the American Taxation Association*, *34*(1), 1–29.

Storper, M., & Venables, A. J. (2004). Buzz: Face-to-face contact and the urban economy. *Journal of Economic Geography*, *4*(4), 351–370.

Tang, T., Mo, P. L. L., & Chan, K. H. (2017). Tax collector or tax avoider? An investigation of intergovernmental agency conflicts. *The Accounting Review*, *92*(2), 247–270.

Wang, J. (2013). The economic impact of special economic zones: Evidence from Chinese municipalities. *Journal of Development Economics*, *101*, 133–147.

Wang, K., & Xiao, X. (2011). Controlling shareholders' tunneling and executive compensation: Evidence from China. *Journal of Accounting and Public Policy*, *30*(1), 89–100.

Wheeler, P. B. (2019). Loan loss accounting and procyclical bank lending: The role of direct regulatory actions. *Journal of Accounting and Economics*, *67*(2–3), 463–495.

Wu, L., Wang, Y., Lin, B.-X., Li, C., & Chen, S. (2007). Local tax rebates, corporate tax burdens, and firm migration: Evidence from China. *Journal of Accounting and Public Policy*, *26*(5), 555–583.
Xu, W., Zeng, Y., & Zhang, J. (2011). Tax enforcement as a corporate governance mechanism: Empirical evidence from China. *Corporate Governance: An International Review*, *19*(1), 25–40.
Zhang, M., Ma, L., Zhang, B., & Yi, Z. (2016). Pyramidal structure, political intervention and firms' tax burden: Evidence from China's local SOEs. *Journal of Corporate Finance*, *36*, 15–25.
Zingales, L. (2017). Towards a political theory of the firm. *Journal of Economic Perspectives*, *31*(3), 113–130.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fan, L., Ouyang, C., Pittman, J., Xiong, J., & Yao, J. (2024). Information acquisition and tax avoidance: Evidence from a natural experiment. *Journal of Business Finance & Accounting*, 1–40. https://doi.org/10.1111/jbfa.12830

APPENDIX A: VARIABLE DEFINITIONS

Variables	Definition
Panel A: Tax avoidance	
ETR	Income tax expenses/pre-tax income (Chen & Lin, 2017; Porcano, 1986)
ETR2	Income tax expense/(pre-tax income − deferred tax expense/statutory tax rate) (Shevlin, 1987)
CETR	Cash income taxes paid/pre-tax income (Bradshaw et al., 2019; Cheng et al., 2012)
CETR2	(Cash income taxes paid)/(pre-tax income − deferred tax expense/statutory tax rate)
TAX_factor	The composite measure of *ETR*, *ETR2*, *CETR*, *CETR2*, *BTD* and *DDBTD* using factor analysis (Lennox et al., 2013). A higher value of *TAX_factor* indicates a higher level of tax avoidance
LRETR	$\sum_{t=-2}^{t} Incometaxexpenses_{it} / \sum_{t=-2}^{t} pre-taxincome_{it}$ (Dyreng et al., 2008)
LRCETR	$\sum_{t=-2}^{t} cashincometaxespaid_{it} / \sum_{t=-2}^{t} Pre-taxincome_{it}$ (Dyreng et al., 2008)
BTD	Total book-tax difference, which is the difference between book income and taxable income scaled by lagged assets (Manzon & Plesko, 2001)
DDBTD	Residual of book-tax difference, which is the residual from the following regression with fixed effects: $BTD_{it} = \beta_1 TA_{it} + \mu_i + \varepsilon_{it}$, where *BTD* is total book-tax difference and *TA* is total accruals (Desai & Dharmapala, 2006)
Panel B: Income shifting	
Shifting	Tax-motivated income-shifting opportunities, which is the product of the difference between the highest and lowest statutory tax rates among group members (i.e., the rate range) and the aggregate amount of related party transactions (scaled by lagged assets of the group) (Lin et al., 2018)
Panel C: High-speed rail	
HSR	An indicator variable with a value of one for year *t* and onward if an HSR service is introduced during year *t* in the city in which a firm is headquartered and zero otherwise
Panel D: The intensity of corporate site visits	
Visit_Inst	Ln (1 + the frequency of site visits that are conducted by institutional investors)
Visit_Anlyst	Ln (1 + the frequency of site visits that are conducted by financial analysts)
Visit_Indiv	Ln (1 + the frequency of site visits that are conducted by individual investors)
Visit_Media	Ln (1 + the frequency of site visits that are conducted by news media)

(Continues)

Panel E: Rent-extraction propensity	
RPT	Abnormal-related party transaction, which is the residual of the following regression model (Jian & Wong, 2010; Jiang et al., 2010; Wang & Xiao, 2011): $RPT\_raw_{it} = \alpha_0 + \alpha_1 RPT\_raw_{it-1} + \alpha_2 ROA_{it} + \alpha_3 TOP1_{it} + \alpha_4 SOE_{it} + \alpha_5 SIZE_{it} + \alpha_6 LEV_{it} + \alpha_7 PPE_{it} + \varepsilon_{it}$ where *RPT_raw* is total related party transactions/total assets
ORECTA	Abnormal other receivables, which is the residual of the following regression model (Jian & Wong, 2010; Jiang et al., 2010; Wang & Xiao, 2011): $ORECTA\_raw_{it} = \alpha_0 + \alpha_1 ORECTA\_raw_{it-1} + \alpha_2 ROA_{it} + \alpha_3 TOP1_{it} + \alpha_4 SOE_{it} + \alpha_5 SIZE_{it} + \alpha_6 LEV_{it} + \alpha_7 PPE_{it} + \varepsilon_{it}$ where *ORECTA_raw* is other receivables/total assets
DIVERGENCE	The difference between control right and cash flow right of controlling shareholder
HHI10	Herfindahl index of shareholdings of the 10 largest shareholders
BUSYDIRECT	The total number of directorships held by independent directors, divided by the number of independent directors (Fich & Shivdasani, 2005)
CEO_OWN	The percentage shareholding of the CEO
BIG10	A dummy variable that equals one if the firm is audited by Top 10 auditors based on an audit firm's Chinese revenues and zero otherwise
Panel F: Control variables	
SIZE	Ln (total assets)
LEV	Total liabilities scaled by total assets
ROE	Pre-tax income/total equity
Q	The sum of market value of tradable shares, book value of non-tradable shares and liabilities scaled by book value of total assets
PPE	Fixed assets/total assets
INTANG	Intangible assets/total assets
INVENTORY	Inventory/total assets
LOSS	An indicator variable that equals one if a firm's net income is negative in the previous year and zero otherwise
BSIZE	Board size, which is the number of board directors
BIND	The ratio of the number of independent directors to the total number of directors
DUAL	A dummy variable, which is equal to 1 if the CEO is also the Chairman
TOP1	The percentage shareholdings of the largest shareholder

(Continues)

SOE	An indicator variable that equals one if the government is the ultimate controlling shareholder and zero otherwise
TREND	It is the post-HSR trend, which equals $t - s_i$ if $t \geq s_i$ and zero otherwise, where s_i denotes the year when the first HSR line opens in the city (Wang, 2013)
LNGDP	The natural logarithm of gross domestic product of the city
LNPEO	The natural logarithm of total population of the city
LNFDI	The natural logarithm of foreign direct investment inflows of the city
DEFICIT	The difference between fiscal expenditure and fiscal revenue scaled by the fiscal revenue
UNEM	The unemployment rate of the city
LNTRSP1	The natural logarithm of freight transport volume of the city
LNTRSP2	The natural logarithm of freight transport volume of the city