Does Digital Technology Development Attenuate Investor Local Attention Bias?[#]

John W. Goodell[a,*], *Mingsheng Li*[b], *Desheng Liu*[c]

[a] *University of Akron, Akron, Ohio 44325*
[b] *Bowling Green State University, Bowling Green, Ohio 43403*
[c] *Qilu University of Technology (Shandong Academy of Sciences), Jinan* 250353, *China*

JEL Classification: G350, G40

*****Corresponding author. E-mail addresses: johngoo@uakron.edu (Goodell), mli@bgsu.edu (Li), and liudeshenggs@126.com (Liu). Authors thank the participants of Financial Seminar Series at Shandong University of Finance & Economics, and Qilu University of Technology for their comments and suggestions.

[#] Authors contribute to the work equally and they are listed alphabetically based on their last names.

Does Digital Technology Development Attenuate Investor Local Attention Bias?

Abstract

The advancement in digital and information technology (DIT) has profound effects on individuals, corporations, and society. The benefits of DIT, such as enhanced efficiency, easily access to information, and increased connectivity, are clear. However, scholars have raised various concerns regarding the plethora of information. For example, how DIT affects investors' information acquisition behavior is unclear. Competing theories offer conflicting predictions in investor information acquisition. These predictions have different implications regarding whether investors acquire more information about local firms (local attention bias), versus about non-local firms, when information becomes easily accessible. Our empirical results show that as DIT develops, investors pay more attention to local firms, amplifying local attention bias. Economic development and a better developed institutional environment amplify rather than attenuate local attention bias. Mediation analysis further shows that DIT development increases attention comovement and stock return correlation not only directly but also indirectly through local attention bias as a mediator. Our novel evidence suggests that when information is more easily accessible associated with DIT development, information asymmetry can be amplified when agents can choose what to learn, increasing polarization of information acquisition and selective exposure to information.

Key Words:
Digital technology; Local attention bias; Attention comovement; Stock return correlation; Emerging markets.

Does Digital Technology Development Attenuate Investor Local Attention Bias?

> "*In a very short period historically, Big Data practices have become normalized and now mediate almost every social interaction, whereas we seem unaware or complacent about their implications on our ways of life*."
> Ngwenyama et al. (2023)

1. Introduction

In an era of the internet and digital world, digital and information technology (DIT) developments have profoundly affected every aspect of our lives and society. The benefits associated with DIT development are well-known and obvious, such as easy access to information, increased connectivity, and enhanced productivity through automation (Birth, 2020; Jiang et al., 2023; Ngwenyama et al., 2023). At the same time, DIT developments also raise growing concerns, such as privacy and security, the reliability and accuracy of information, and undermining of human freedom and dignity (DeMoya and Pallud, 2020). Ngwenyama et al. (2023) warn that digital technology can have grave social consequences and even "*can condition our reality and entrap us into an open prison.*"

In this study, we engage in this academic discussion and shed additional light on the literature by investigating how DIT affects investors' information acquisition, which has profound effects on capital allocation and stock market efficiency. On one hand, DIT developments, such as big data, machine learning, cloud computing and artificial intelligence (AI), make information more accessible to all investors at low or no cost (Dugast and Foucault, 2018; Huang et al., 2021). On the other hand, DIT development can lead to information overload (Agrwal et al., 2017; Mendelson and Pillai, 1998) because individuals not only have limited capacity to process information (Golman et al., 2017; Hirst and Hopkins, 1998; Shiffrin and Schneider, 1977) but also have limited attention to information (Golman et al., 2017; Peng, 2005). Levy (2021) further shows that online information and social media limit individuals' exposure to counter-attitudinal news and increase polarization because of limited attention and the plethora of information.

An emerging phenomenon related to investor attention to information and information acquisition is local attention bias. Namely, investors pay more attention to and acquire more information about local firms than non-local firms (Huang, Qiu and Wu, 2016; Dyer, 2021). The underlying theoretical explanation for local attention bias is based on information endowment, as proposed by Van Nieuwerburgh and Veldkamp (2009). Specifically, Van Nieuwerburgh and Veldkamp (2009) argue that investors are endowed with an initial information advantage for their local firms and tend to acquire more information about these firms. Cziraki, Mondria and Wu (2021) develop a similar conceptual framework. Broadly, local attention bias is related to other behavioral biases, such as selective exposure and attention, confirmation bias (Festinger, 1957; Sims, 2003; Knobloch-Westerwick, 2014; Andrei and Hasler, 2020), selective inattention and information avoidance, or the '*ostrich effect*' (Karlsson et al., 2009; Sicherman et al., 2016; Golman, Hagmann, and Loewenstein, 2017).[1]

Given these insightful studies, an important but less explored issue is whether DIT development reduces or increases investors' advantage in information endowment, as suggested by Van Nieuwerburgh and Veldkamp (2009), consequently attenuating or amplifying local attention bias. Intuitively, DIT developments provide geographically distant investors equal access to information about non-local firms, reducing local investors' advantage in information endowment and attenuating local attention bias. However, at a deeper level, opposed factors affect

[1] Although it is not the focus of the current paper, local attention bias is also related to the well-documented local investment puzzle (Coval and Moskowitz, 1999; Ivkovic and Weisbenner, 2005). For example, Peress (2004), Van Nieuwerburgh and Veldkamp (2010), and Cookson, Engelberg, and Mullins (2023) argue that local investors are incentivized to acquire information about the local firms of their portfolios because they tend to hold more stocks of local firms in their portfolio than stocks of non-local firms.

investor information acquisition either as strategic substitutes or strategic complements, which will result in opposite effects on local attention bias.

On the one hand, the conventional wisdom proposed by Grossman and Stiglitz (1980) reveals that as more agents acquire information about a firm, the stock price becomes more informative (also see Goldstein and Yang, 2015; Veldkamp, 2006a, 2006b). Therefore, other agents freeride on the market price, which reduces agents' incentive for information collection, resulting in a strategic substitute in learning and information acquisition. Thus, in the context of local attention bias, the substitute effect predicts either a decline or no change in the local attention bias because information for both local and non-local firms can be more easily collected due to technological development. We refer to this as a *substitute effect hypothesis*.

On the other hand, Barlevy and Veronesi (2000) show that as more agents acquire information, prices may not necessarily become more informative, which generates greater incentives for more agents to acquire information, resulting in strategic complementarity in information acquisition. Similarly, Banerjee et al. (2019) indicate that improving access to information about asset fundamentals can be counterproductive when speculative motivation dominates. This is because agents have incentives to learn more about both the asset fundamental and others' beliefs. Veldkamp (2006a, 2006b) provides different insights based on information production and supply and indicates that information acquisition could be strategically complementary. Thus, strategic complementarity in information acquisition is expected to amplify local attention bias. We refer to this as *a strategic complementary effect*.

Empirically, how DIT development affects local attention bias is as ambiguous as the theoretical predictions. A way to describe the situation is the quote of Hellwig and Veldkamp (2009), who state *"agents who want to do what others do, want to know what others know (p223)*."

Conversely, it is also true that agents want to know that others do not know if they want to do what others don't do. Hellwig and Veldkamp (2009) further propose that the information people observe depends both on its availability and on their choice of what to learn. Predictions by various models reflect the authors' assumptions rather than investors' actual behavior. Moreover, Goldstein and Yang (2015) demonstrate that information acquisition can be either strategic complementary or substitute, depending on whether the value of acquiring information about other fundamental decreases or increases as more traders become informed about one fundamental. Thus, how DIT development affects local attention bias is a pure empirical question.

We strive to shed light on literature by providing empirical evidence in a large emerging market. Using 10,546 firm-year observations of publicly listed Chinese A-share companies from 2011 to 2020 and employing a comprehensive index, *Digit,* as a proxy for DIT development, we find that DIT development significantly magnifies investor local attention bias. The result is consistent based on a battery of robustness tests, such as controlling for firm characteristics, local economic environments, industry-, year-, and location-fixed effects. In terms of economic significance, an increase in one standard deviation of *Digit* increases location attention bias by approximately 1.9% based on the mean value of local attention bias proxied by internet search volume. We also use two different interaction variables as instrumental variables (IVs) for *Digit* and conduct two-stage least squares (2SLS) regressions. The results based on the IV approach confirm our baseline regression, suggesting that the positive relation is unlikely due to omitted variables.

Next, we investigate geographic heterogeneity by considering three factors. First, we classify firms into two subgroups based on the economic development of the province where a firm's headquarters is located. We consider this factor because economic and technological developments in a region enhance each other and affect local people's economic wellbeing and decision making. We find that the positive effect of *Digit* on local attention bias exists only for firms located in provinces with higher economic development measured by GDP per capita. This suggests that economic development amplifies rather than attenuates local attention bias.

The second factor that we consider is the institutional environment. We classify firms into two subgroups based on a marketization index (MI) developed by Wang et al. (2018). Firms in the MI-High (MI-Low) subgroup face a better (less) developed institutional environment. Our conjecture is that firms headquartered in a place with a better institutional environment face less political uncertainty and are more transparent, attracting more attention from investors, both local and non-local, ceteris paribus, which may attenuate local attention bias. However, our results show that the positive effect of *Digit* on local attention bias mainly exists for the MI-High subgroup. This indicates that a better developed institutional environment amplifies local attention bias.

The third factor we consider is the evolution of technological development. Specifically, we divide the whole sample period into two subperiods based on the average value of *Digit* of all provinces in each year. The later subperiod (after 2016) is classified as Digit-High because the mean value of *Digit* is larger than that in the earlier subperiod (before and including 2016). Our results indicate that the positive effect mainly exists in the Digit-High subperiod. In relative terms, based on the magnitude of regression coefficients, the positive effect of *Digit* on local attention bias in the Digit-High subperiod is approximately 1.35 times larger than that in the Digit-Low period when other control variables are not included or approximately 2.65 times larger when all other control variables are included. This additional evidence not only provides more nuanced information but also further confirms that the positive effect of *Digit* becomes stronger as DIT develops.

To shed additional light on the literature, we continue to investigate whether local attention bias influences an emerging phenomenon of attention comovement among investors (Drake et al., 2017). This extension is important because Drake et al. (2017) show that attention comovement is a primary cause for excess stock return correlation (Barberies, Shleifer, and Wurgler, 2005; Hirshleifer, 2015). However, limited evidence exists on whether local attention bias affects attention comovement and stock return correlation. Our results show that local attention bias positively influences attention comovement. In terms of economic significance, a one standard deviation increase in attention bias is associated with an approximately 16.5% to 20.6% increase in attention comovement when attention comovement is measured by internet search volume. Additionally, we find that an increase of one standard deviation in local attention bias increases stock return comovement by approximately 17.7% to 24.4%, depending on model specifications.

After showing that *Digit* positively affects local attention bias, which in turn positively influences attention comovement and stock correlation, we take a further step and investigate whether *Digit* also affects attention comovement and stock correlation. If so, is the influence direct or indirect through local attention bias as a mediator? Employing a widely used mediation analysis method in the literature (Baron and Kenny, 1986; Wen and Ye, 2014), we find that *Digit* affects attention comovement and stock return correlation both directly and indirectly through local attention bias as a mediator. Specifically, for the direct effect, an increase in *Digit* by one standard deviation is associated with an increase in attention comovement by approximately 11.0% based on the absolute mean value of attention comovement measured by internet search volume, and the increase in stock return correlation is approximately 6.5%. The indirect effect through location attention as a mediator is approximately 7.5% of the direct effect. Moreover, the effect of local attention bias on attention comovement remains positive and significant after controlling for *Digit* and other control variables.

The current paper contributes to literature in several ways. First, to the best of our knowledge, no research has attempted to investigate whether digital and information technology development affects local attention bias. This omission is unfortunate because we are living in an information and digital era, and local attention bias has been found to be associated with biased behavior in information acquisition and capital allocation. Additionally, this is a timely issue to investigate because scholars have raised alarming concerns regarding social harm emerging from big data and diffusion of online information (DeMoya and Pallud, 2020; Ngwenyama et al., 2023). Second, although scholars have discovered attention comovement among investors and excess stock return correlation across firms, this is the first study showing that local attention bias is one of the causes for attention comovement and stock return correlation. This discovery is important since information acquisition fundamentally affects price informativeness and capital market efficiency (Bond et al., 2012; Gross and Stiglitz, 1980; Verrecchia, 1980). Third, our nuanced mediation analysis and empirical findings support Van Nieuwerburgh and Veldkamp's (2009) theoretical prediction that information advantages are not only sustainable when information is mobile but also that information asymmetry can be amplified when investors can choose what to learn. Fourth, our study is also related to an emerging literature showing that social media increases polarization in attention (Levy, 2021) or amplifies echo chambers (Cookson et al., 2023) and that improving access to information and the advancement of speed technology may negatively affect price informativeness (Banerjee et al., 2018; Dugast and Foucault, 2018; Huang and Yueshen, 2021). Finally, the findings in this paper have important implications for scholars, regulators, and policy makers, especially in emerging markets, because these markets experience excess stock return comovement and high price synchronicity due to investors' herding behavior and less mature financial markets (Li et al., 2020; Morck, Yeung, Yu, 2000).

2. Brief literature review and hypothesis development

A. Literature on selective exposure and attention (inattention)

Selective exposure and attention have been widely documented in many fields, such as psychology, behavioral studies, and social and communication research. Knobloch-Westerwick (2014) indicates that information choice is endogenous because people do not just receive messages but choose among them. According to cognitive dissonance theory (Festinger, 1957), individuals seek and acquire information aligned with preexisting attitudes, resulting in confirmatory information acquisition (Scherer et al., 2013) and selective exposure (Peress, 2004; Van Nieuwerburgh and Veldkamp, 2009; 2010). On the other hand, individuals also actively attempt to reduce dissonance by avoiding circumstances that likely induce dissonance, leading to information avoidance and rational inattention (Golman et al., 2017; Sims, 2003). An alternative explanation for selective attention is that attention is a limited resource and is selectively employed to facilitate information processing (Golman et al., 2017; Shiffrin and Schneider, 1977). Hirst and Hopkins (1998) and Peng (2005) further indicate that the time and attention needed to process financial information is nontrivial and that investors have limited time and attention to process information. Other studies further find that individuals are intentionally exposed to more polarized information (Peress, 2004; Van Nieuwerburgh and Veldkamp, 2010).

An emerging finance literature provides pervasive empirical evidence on investor selective attention and exposure or information confirmation behavior. For example, Cookson et al. (2023) use more than 400,000 users of Stock Twist, a social network of investors, and find that investors deliberately choose to consume information that aligns with their prior views, a phenomenon also known as echo cambers that serves as a mechanism of sustaining disagreement. Specifically, Cookson et al. (2023) show that self-described bullish investors are five times more likely to follow a user with a bullish view of the same stock than are self-described bearish investors. More importantly, they find that selective attention and exposure exist not only among nonprofessionals and active users of Stock Twist but also among professionals and less active users. Karlsson, Loewenstein and Seppi (2009) show that individuals actively monitor their portfolios when the stock market is performing well but tend to avoid looking at their portfolio performance when markets are down. This is commonly referred to as the "ostrich" effect. Ehrmann and Jansen (2022) further find that investor inattention causes excess stock return comovement across firms.

B. Literature on local attention bias

Given the well-known phenomenon of selective attention and exposure as a behavioral-related bias, it is not a surprise to observe local attention bias. For example, Huang et al. (2016) find that individual investors spend much more time analyzing the stocks of firms whose headquarters are close to their own geographical locations. Specifically, they show that the average percentage of internet posts on a firm from posters in the same province as the firm is approximately 9.75%, which is about 94.6% (9.75/5.01 - 1) higher than the average percentage of internet users in the province (5.01%). Local attention bias also exists among institutional investors. For example, Dyer (2021) shows that institutional investors acquire approximately 20% more financial information for their investment in local firms than in nonlocal firms. Their analysis further indicates that the observed local attention bias is due to a combination of several factors, such as behavioral biases, industry affiliation, local familiarity, and portfolio holdings.

A mainstream theoretical explanation of local attention bias is the information endowment advantage proposed by Van Nieuwerburgh and Veldkamp (2009; 2010). Specifically, these authors argue and show that the prior information that local investors have about local assets' payoff is slightly more precise than the prior information that nonlocal investors have. The initial information advantage possibly reflects what one incidentally observes about the local environment. Consequently, local investors endowed with a small home-information advantage choose to learn more about local firms because they profit more from knowing information that others do not know. Thus, the interaction of information acquisition and portfolio investment decisions causes local investors to acquire information that magnifies their comparative advantage in home assets. They further suggest that persistence in information immobility is possible not because investors cannot learn what locals know or because such information is expensive, but because investors do not choose to learn what others know. Cziraki et al. (2021) provide a similar framework and show that investors tend to acquire more information about an asset when the investor has a small initial information advantage about the asset.

C. Information acquisition and hypothesis development

A deeper fundamental question is what factors affect investors' information acquisition. Two streams of literature provide different explanations. One explanation rests on strategic substitution in information acquisition, and another one is based on strategic complement. The strategic substitute theory traces back to Grossman and Stiglitz (1980), who show that as more agents acquire information about a firm, the stock price becomes more informative. Therefore, other agents have less incentive to acquire information because they can infer private information from the market price, resulting in a strategic substitute in learning and information acquisition. Similarly, Goldstein and Yang (2015) argue and show that the strategic substitution effect occurs when an increase in the mass of agents acquiring information on one fundamental leads fewer agents to acquire information on the other fundamentals because of the "inference" effect. They further note that information acquisition could also be a strategic complement if the uncertainty effect dominates the inference effect. However, learning the same information is always a strategic substitute as more traders become informed about the fundamental, and the value of acquiring the information about the fundamental decreases. Hellwig and Veldkamp (2009) further show that the substitute effect occurs when agents prefer to differentiate their information choices from others.

Applying the substitute effect theory of information acquisition to the context of how DIT development affects local attention bias, we expect either a decline or no change in local attention bias because the information for local firms is equally available to both local and nonlocal investors as DIT develops, reducing the incentive for local agents to collect local information in the spirit of Grossman and Stiglitz (1980) and reducing or eliminating the information endowment advantage in the spirit of Van Nieuwerburgh and Veldkamp (2009). Accordingly, we propose the following *substitute effect* hypothesis:

Hypothesis 1: *Local attention bias declines as digital and information technology develops.*

Additionally, the substitute effect hypothesis is also supported by both utility theory and standard intuitions because when uncertainty is high, it is more valuable to acquire information (Kendall, 2018; Lewis, 1999). For example, Kendall (2018) finds that traders are willing to pay the most for information when uncertainty is highest. Similarly, Andrei and Hasler (2020) show that greater uncertainty increases investors' attention to news because greater uncertainty is associated with greater volatility of expected returns, which increases the likelihood of large future trends. Thus, combining the utility of uncertainty reduction with the fact that local investors have information endowment for local firms, it is natural to expect that local investors tend to acquire more information about nonlocal firms that are more uncertain to them, especially when information for nonlocal firms becomes more accessible with DIT development.

Another stream of literature suggests that information acquisition is equally likely to be strategic complementary. For example, Barlevy and Veronesi (2000) argue and show that prices may not necessarily become more informative as more agents acquire information, which generates greater incentives for more agents to acquire information, resulting in strategic complementarity. This is in sharp contrast with strategic substitute theory, which assumes that prices become more informative when more agents acquire information. Barlevy and Veronesi's (2000) justification for the strategic complement theory is that an increase in the fraction of informed traders can cause prices to be more extreme, and it makes it harder for investors to infer the true signal when prices are extreme. Similarly, Banerjee et al. (2019) show that improving access to information about asset fundamentals can make price less informative when speculative motivation dominates because agents have incentives to learn more about both the asset fundamental and others' beliefs. Additionally, they posit that greater public disclosure can crowd out private learning about fundamentals and encourage more learning about others, creating learning complementarity. Moreover, Goldstein and Yang (2015) demonstrate that information acquisition can be strategic complementary because the increase in the mass of agents acquiring information on one fundamental can reduce the uncertainty of the fundamental and cause more agents to acquire information on other fundamentals.

Veldkamp (2006a) takes a different approach and shows that information acquisition is strategically complementary. Specifically, this author introduces markets for information production and argues that competitive producers of information charge more for low-demand information than for high-demand information because information production has high fixed costs. Investors want to purchase the same information that others purchase because high-demand information costs less. Additionally, Veldkamp (2006a) argues that information is a nonrival good with a fixed cost of discovery and a low marginal cost of replication. Thus, the decreasing price in quantity generates complementarity in information acquisition. Veldkamp (2006b) further argues that information acquisition complementarities lie not in the demand for assets but in the demand for information used to price the asset. Accordingly, we apply the strategic complement theory of information acquisition to the context of how DIT development affects local attention bias and posit the following complementary effect hypothesis:

Hypothesis 2: *Local attention bias increases as DIT develops due to strategic complementarity in information acquisition.*

3. Variable descriptions, data, and sample selection

A. Measure for local attention bias

The primary dependent variable used in our study is investor local attention bias (*LoAtt*). We follow the mainstream literature and measure investors' attention on a firm using internet search volume for the firm. In the U.S., scholars use Google searches for the firm's stock ticker (Cziraki et al., 2021; Da, Engelberg and Gao, 2011; Drake et al., 2017; Huang, Huang, and Lin, 2019). We use Baidu search volume for A-share firms' stock tickers in China because Baidu is the most widely used search engine in China, similar to Google in the U.S. Accordingly, we follow Cheng and Lu (2019) and Huang, Qiu, and Wu (2016) and compute firm *i*'s local attention in year *t* as follows:

$$LoAtt_{i,j,t} = Ln[1+(S_{i,j,t}/S_{i,t})/(IAP_{j,t}/IAP_t)]. \quad (1)$$

Where $S_{i,j,t}$ is the summation of daily internet (Baidu) searches for *firm i* headquartered in province *j* by people in province *j* in year *t*. $S_{i,t}$ is the summation of daily internet searches for *firm i* by people in all provinces in year *t*. $IAP_{j,t}$ is the number of internet access ports in province *j* in which firm *i* is headquartered in year *t*, and IAP_t is the number of internet access ports in all provinces in year t. Intuitively, $S_{i,j,t}/S_{i,t}$ scaled by $IAP_{j,t}/IAP_t$ measures the proportion of local internet searches (local attention) relative to the proportion of local internet accessibility with regard to the whole country.

As an alternative measure of local attention bias, we compute $LoAttx_{i,t}$ as $Ln\ [1+(S_{i,j,t}/S_{i,t})/(IU_{j,t}/IU_t)]$, where $IU_{j,t}$ and IU_t are the number of internet users in province j in which firm i is headquartered and in all provinces, respectively, in year t (Huang et al., 2016). Obviously, $IU_{j,t}/IU_t$ provides a more accurate measure of the proportion of internet users in province j relative to the whole country than $IAP_{j,t}/IAP_t$ does. However, the data for the number of internet users at the province level are not available after 2016. Thus, we employ *LoAtt* as our primary measure of local attention bias and use *LoAttx* only for robustness test purposes. The data for the Baidu daily search volume are downloaded from Baidu.com for each firm using its ticker, and the IAP data are retrieved from the *China Urban Statistics Yearbook* for each province in each year.

Table 1 provides more detailed information about variable definitions. Table 2 reports descriptive statistics of the main variables. The mean and median values of *LoAtt* are 1.972 and 1.906, respectively, ranging from 0.0 (minimum) to 4.349 (maximum).[2] The large range and standard deviation (0.503) indicate a large variation in local attention bias. The statistical distribution based on the alternative measure of *LoAttx* shows a similar pattern, with a mean (median) value of 1.837 (1.811), ranging from 0.493 to 4.032.

[Insert Tables 1 and 2 about here]

B. Measure for digital and information technology development

Literature does not offer a standardized way of measuring digital and information technology development. Thus, we follow a few seminar papers that investigate digital finance

[2] The minimum value of zero for *LoAtt* is due the result of one company (with a stock ticker of 002653) headquartered in Xi Zang (Tibet) whose Baidu search volume by the local people was zero in 2019.

and the digital economy in China and employ a digitalization (*Digit*) index in the province where a firm's headquarters is located as a proxy for the level of technological development in the local area (Guo et al., 2020; Zhang et al., 2020). This index is a comprehensive measure, and it considers a wide range of factors. Specifically, $Digit_{i,j,t}$, the digitalization index in province j in which firm i is headquartered in year t, is computed as the principal component of the following five groups of variables in the province: (1) the number of internet users per 10,000 population; (2) the percentage of the labor force specializing in the internet and computer service sectors; (3) the industrial output of the telecommunication and information technology sectors; (4) the percentage of people using mobile phones; and (5) the digital financial inclusion index developed by the Research Center of Digital Finance of Beijing University (Guo et al., 2020; Zhao et al., 2020). A higher value of *Digit* indicates that DIT is better developed in the province, ceteris paribus. The data for *Digit* are collected from the *China Urban Statistics Yearbook.* The mean and median values of *Digit* are 0.197 and 0.115, respectively, ranging from 0.004 to 0.818 (Table 2).

C. Sample selection

We choose the period from 2011 to 2020 as our sample period because Baidu has imposed strict restrictions on how much search data can be downloaded by one registered account after 2020. This restriction on data downloading makes it practically impossible for us to download the daily search volume of all publicly listed companies. Therefore, for the sample composition, we include all publicly traded A-share companies listed on the main stock exchanges in China between 2011 and 2020 in our initial sample mainly because of unavailability of Baidu search volume data. We retrieve firms' financial and corporate governance data from the GTA dataset and supplement missing data from the RESSET and CCER databases. Then, we follow the literature and filter the initial sample by excluding firms in finance, banking, real estate, insurance industries and firms with missing financial data and nonnormal trading status of ST, ST*, or PT. The total number of firm-year observations in the final sample is 10,546.

4. Test method and empirical results

A. Baseline regression

We conduct the following baseline regression to investigate the effect of DIT development on investor local attention bias:

$$LoAtt_{i,j,t} = \alpha_{0,i} + \alpha_{1,i} \times Digit_{i,j,t} + \boldsymbol{\Phi} \times \boldsymbol{FirmC} + \boldsymbol{\lambda} \times \boldsymbol{LocalC} + \boldsymbol{FE}(Year/Ind/Location) + \epsilon_i. \quad (2)$$

Where $LoAtt_{i,j,t}$ and $Digit_{i,j,t}$ are the proxies for location attention bias of firm *i* and DIT development in province *j* where the firm's headquarters is located in year *t*, respectively. ***FirmC*** includes two sets of variables controlling for firm characteristics. The first set of variables controls for firm size (*Ln(1+total assets)*), capital structure (*Leverage*), return on assets (*ROA*), growth potential (*SaleGrowth*), book-to-market ratio (*BK/Mkt*), stock performance (*Return*), and state ownership (*SOE*). The second set of variables measures a firm's publicity and stock market liquidity, namely, the number of analysts covering the stock (*Analyst*), whether the firm's stock is included in the major stock index (*HS300-index*), institutional investor ownership (*InstSH(%)*), and annual stock trading volume (*Ln(TrdVol)*). ***LocalC*** is a set of variables controlling for local economic development (*GDP/P*), local educational level (*CollegeP(%)*), local demographics (*PopGrowth(%)*), and whether the firm's headquarters is located in a central city or not (*CenterCity*).

The baseline regression results are reported in Table 3. In Model (1), we include only *Digit* as an independent variable and control for year-, industry-, and location-fixed effects. The year fixed effects control for changes over time across all provinces. The location fixed effects control for all time-invariant differences across provinces. The industry fixed effects control for industryspecific effects within each industry. The coefficient on *Digit* is 0.181 and statistically significant at the 0.01 level. In terms of economic significance, an increase of one standard deviation (0.207) in *Digit* is associated with an increase in local attention bias of 0.037 (=0.181x0.207), equivalent to a 1.9% (=0.037/1.972) increase based on the mean value of *LoAtt*, which is 1.972. Additionally, the adjusted R^2 is 0.680, indicating that approximately 68% of variables in location attention bias can be explained by the variation of *Digit* after controlling for fixed effects. In models (2) and (3), we include firm control variables in two subgroups, and the coefficients on *Digit* are 0.173 and 0.160, with both numbers being statistically significant at the 0.01 level. The adjusted R^2s are 0.696 and 0.737, respectively, and the differences are relatively small compared with that in model 1. In model (4), we include control variables of both firm characteristics and local factors. The coefficient on *Digit* is 0.158, remaining statistically significant at the 0.01 level. In terms of economic significance, an increase in one standard deviation in *Digit* is associated with a 1.7% (=0.158 x 0.207/1.972) increase in local attention bias based on the mean value of *LoAtt*. These results provide a strong support for our complementary effect hypothesis, but rejecting the substitute effect hypothesis.

[Insert Table 3 about here]

B. *2SLS regressions with different instrumental variables*

Although we control for firm characteristics and local environments along with various fixed effects (industry, year, and location) in our baseline regressions, there is still a possibility that the observed positive effect of *Digit* on local attention bias is caused by omitted variables. Thus, we conduct two-stage least squares (2SLS) regressions using two different instrumental variables (IVs). Specifically, we follow the identification strategy of Nunn and Qian (2014) and use $Ph_{84} \times IU_{t-1}$ as the first IV. Ph_{84} is the number of landline phone users per ten thousand persons in a province in 1984, a time-invariant proxy of the province's historical telecommunication technology.[3] IU_{t-1} is the lagged number of internet users in the whole country in year *t*-1, a time-variant proxy for overall telecommunication development. The advantage of this identification strategy is that $Ph_{84}\times IU_{t-1}$ varies by province and time, allowing for further control for year and location fixed effects. Additionally, this IV is expected to affect *Digit* but not *LoAtt* except through *Digit*. Our second IV is $RDLS\times IU_{t-1}$, where *RDLS* is the relief degree of the land surface in a province, an alternative time-invariant proxy for the province's telecommunication technology. This use is justified because regions with a large value of *RDLS* are hilly areas and have unfavorable geographic environments for the development of telecommunication technologies (Feng et al., 2023).

Table 4 reports the results. In the first-stage regressions of the 2SLS models (Columns 1 and 3), the dependent variable is $Digit_{i,t}$, and independent variables include IV (=$Ph_{84}\times IU_{t-1}$ or $RDLS\times IU_{t-1}$) and all other control variables as indicated in the baseline regression. The coefficients on the IV are 0.021 and 0.008 in columns 1 and 3, respectively, and both numbers are significant at the 0.01 level. This indicates that these IVs positively affect *Digit*. In the second-stage regressions (Columns 2 and 4), the primary independent variable is the instrumented *Digit*, *Digit(IVed1)* and *Digit*(*IVed2*), which are the estimated values of *Digit* from the first-stage regressions. The coefficients on these instrumented IVs are 1.573 and 1.479 in columns 2 and 4, respectively, and both numbers are significant at the 0.01 level. These results are consistent with the baseline results, indicating that the positive effect of *Digit* on local attention bias is unlikely

[3] We use the information in 1984 arbitrarily. Our main consideration is that the number of landline phone users in a region many years ago captures the historical development of tele-communication technology in the region, but it is not expected to affect current investors' local bias for a particular firm today. Thus, this variable helps control for region-specific variables associated with telecommunication technology. As a further verification, we also use the number of landline phone users in other years such as 1980 and 1985, which does not affect our results.

due to omitted variables. Additionally, the Kleibergen–Paap rk LM statistics are 2493 and 461, and Stock-Yogo weak ID tests are 3968 and 887 in columns 2 and 4, respectively. These statistics verify the validity of the IVs.

[Insert Table 4 about here]

C. *Alternative measure of local attention bias*

As a last robustness test, we use an alternative measure of local attention bias $LoAttx_{i,t}$, which is computed as $Ln\ [1+(S_{i,j,t}/S_{i,t})/(IU_{j,t}/IU_t)]$, where $IU_{j,t}$ and IU_t are the number of internet users in province *j* in which firm *i* is headquartered and in all provinces, respectively, in year t (Huang et al., 2016). However, the data for IU at the provincial level are unavailable after 2016. Thus, the test using *LoAttx* as an alternative measure of local attention bias is based on a subsample from 2011 to 2016. Table 5 reports the results. In column 1, *Digit* is the only independent variable, and its coefficient is 0.167, which is significant at the 0.01 level. In columns 2 to 4, we gradually include other control variables, and the coefficients on *Digit* are 0.143, 0.153, and 0.136, respectively. All these numbers are statistically significant at the 0.01 level. This confirms our results reported earlier, indicating that DIT development positively affects local attention bias.

[Insert Table 5 about here]

D. Considering location heterogeneity and change in DIT development over time

To provide more nuanced results on how geographic factors influence the effect of digital technology on local attention bias, we conduct the following three tests. First, we consider the economic development of the province where the firm's headquarters is located because the economic environment affects not only technology development but also people's behavior. We class a firm as undeveloped (developed) if the GPD per capita of the province in which the firm is headquartered is below (equal to or above) the mean value of GDP per capita of all provinces in year *t*. Panel A of Table 6 reports the results. For firms in the underdeveloped subgroup, the coefficient on *Digit* is 0.008 and statistically insignificant at the 0.1 level in column 1 when other control variables are not included. The coefficient on *Digit* (-0.065) remains statistically insignificant at the 0.1 level when all other control variables are included in column 2. For firms in the developed subgroup, the coefficients on *Digit* are 0.291 and 0.223, and both numbers are statistically significant at the 0.01 level, regardless of whether other control variables are included (column 4) or not (column 3).

[Insert Table 6 about here]

In the second test, we consider the formal institutional environment because intuitively, the institutional environment in an area influences both the local area's economic and technological developments. We classify firms into two subgroups based on a comprehensive marketization index (MI) at the provincial level in China compiled by Wang et al. (2018). Firms are classified as MI-High (MI-Low) if the MI index in the province of the firm's headquarters is greater (equal to or smaller) than the mean value of the MI index of all provinces. Firms in the MI-High subgroup have better developed institutional environments than firms in the MI=Low subgroup. We conduct the baseline regression separately for the MI-Low and MI-High subsamples and report the results in Panel B of Table 6. For the MI-Low subgroup, the coefficients on *Digit* are 0.023 (statistically insignificant at the 0.1 level) and 0.085 (statistically significant at the 0.1 level) in columns 1 and 2, respectively. For the MI-High subgroup, the coefficients on *Digit* are 0.342 and 0.275, which are not only larger in magnitude but also statistically stronger (significant at the 0.01 level) than those for the MI-Low subgroup.

In the third test, we divide the firms into two sub-periods based on the digitalization trend over the sample period. In each year, we compute the mean value of *Digit* of provinces and plot it over time. The unreported result shows that the mean value of *Digit* is larger after 2016 than it is before 2016.[4] Thus, we classify observations before 2016 (including 2016) as Digit-Low and observations from 2017 to 2020 as Digit-High. Panel C of Table 6 reports the results. For the Digit-Low subgroup, the coefficients on *Digit* are 0.09 and 0.069 in columns 1 and 2, respectively, and both numbers are statistically insignificant at the 0.1 level. For the Digit-High subgroup, the coefficients on *Digit* are 0.212 and 0.252 in columns 3 and 4, respectively, which are significant at the 0.01 level. In relative terms of magnitude, the positive effect in the Digit-High subperiod is approximately 1.35 (0.212/0.090 - 1) to 2.65 times (0.252/0.069 -1) larger than that in the Digit-low period. Consistent with the baseline results, this additional evidence indicates that DIT development positively affects local attention bias, and the effect becomes stronger as DIT develops.

5. Mediation analysis of attention comovement and stock return correlation

A large body of literature has shown the presence of investor attention comovement (Drake et al., 2017), which is one of the primary causes of stock return comovement (Barberies, Shleifer, and Wurgler, 2005; Hirshleifer, 2015). However, limited evidence exists on whether local attention bias affects these comovements. If so, whether it is caused by DIT development through local attention bias as a mediator or DIT development also directly affects these comovements. Thus, we strive to fill the gap and investigate these issues as a natural extension of our baseline test. To carry out these tests, we first follow Brake et al. (2017) and estimate attention comovement by conducting the following regression for each firm in each year:

$$FirmAttention_{i,w} = ß_0 + ß_1 IndustryAttention + ß_2 MarketAttention + \epsilon_{i,w}, \tag{3}$$

[4] Due to limited space, the chart showing the digitalization trend is not reported in the manuscript, but available from authors upon request.

Where *i* and *w* index firms and weeks. *FirmAttention* is measured in two different ways: internet (Baidu) search volume and the number of forecast updates by financial analysts who cover the firm each week. *IndustryAttention* is the equal-weighed attention for all firms (excluding firm i) in firm *i*'s industry for a given week based on two-digit industry code classification. *MarketAttention* is computed as the equal-weighted total attention of all firms (excluding firm i) in each week. Then, we use the R^2 estimated by Eq. (3) to compute the firm's attention comovement in year *t* as follows:

$$Att\text{-}Co_{i,t} = Log\ (R^2/(1-R^2)) \tag{4}$$

We also compute stock return comovement (*Return-Co*) using Eq. (3) and Eq. (4) by replacing attention measures with weekly returns (Morck, Yeung, and Yu, 2000). Then, we use the following regression to investigate the effect of local attention bias on attention comovement and stock return correlation:

$$Att\text{-}Co_{i,t} = ß_{0,i} + ß_{1,i} \times LoAtt_{i,t} + \boldsymbol{\Theta} \times \boldsymbol{OtherControls} + \eta_{i,1}. \tag{5}$$

Where ***OtherControls*** refers to the control variables specified in the baseline regression Eq. (2).

Table 7 reports the results. In Columns 1 and 2, the dependent variable is attention comovement measured by Baidu search volume (*Att-Co-SV*). The coefficient on *LoAtt* is 0.103 and significant at the 0.01 level when other control variables are not included (column 1). In terms of economic significance, if location attention bias increases by one standard deviation (0.503), the attention comovement increases by 0.05 (0.103×0.503), or approximately 16.5% (0.103×0.503/0.320) based on the absolute mean value (0.320) of *Att-Co-SV*. The coefficient on *LoAtt* is 0.131 in column 2 and significant at the 0.01 level with all other control variables included (in column 2). Thus, an increase of one standard deviation in local attention bias increases attention comovement by approximately 20.6% (=0.131×0.503/0.320) based on the absolute mean value of attention comovement.

[Insert Table 7 about here]

In Columns 3 and 4 of Table 7, the dependent variable is attention comovement measured by analyst forecast update (*Att-Co-FA*). The coefficients on *LoAtt* are 0.089 and 0.099, which are statistically significant at the 0.1 level. In terms of economic significance, an increase of one standard deviation (0.503) in local attention bias increases *Att-Co-FA* by approximately 2.3% (= 0.099 ×0.503/2.132) based on the absolute mean value of *Att-Co-FA* (2.132). In Columns 5 and 6, the dependent variable is stock return comovement (*Return-Co*), the coefficients on *LoAtt* are 0.178 and 0.129, respectively, and both numbers are significant at the 0.01 level. In terms of economic significance, an increase of one standard deviation in local attention bias increases *Return-Co* by approximately 17.7% (= 0.129 ×0.503/0.367) to 24.4% (= 0.178 ×0.503/0.367) based on the absolute mean value of *Return-Co* (0.367). This evidence shows that local attention bias positively affects attention comovement and stock return correlation across firms.

To investigate whether DIT development directly affects attention comovement and stock return correlation or indirectly through local attention bias acts as a mediator, we follow Baron and Kenny (1986) and Wen and Ye (2014) and conduct the following mediation analysis:

$$Att\text{-}Co_{i,j,t} = c_{0,i} + c_{1,i} \times Digit_{i,j,t} + \boldsymbol{\Theta} \times \boldsymbol{OtherControls} + \epsilon_{i,1}. \quad (5A)$$

$$LoAtt_{i,j,t} = \alpha_{0,i} + \alpha_{1,i} \times Digit_{i,j,t} + \boldsymbol{\Phi} \times \boldsymbol{OtherContorls} + \epsilon_{i,2}. \quad (5B)$$

$$Att\text{-}Co_{i,j,t} = c_{0,i} + c'_{1,i} \times Digit_{i,j,t} + b_{1,i} \times LoAtt_{i,j,t} + \boldsymbol{\lambda} \times \boldsymbol{OtherControls} + \epsilon_{i,2}. \quad (5C)$$

In the context of Wen and Ye's (2014) framework, Eq. (5A) investigates the total effect of the independent variable (*X*=*Digit*) on the dependent variable (*Y*=*Att-Co*). Thus, the coefficient $\boldsymbol{c_{1,i}}$ in Eq. (5A) captures the total effect of DIT development on attention comovement. Eq. (5B) investigates the effects of the independent variable (*X=Digit*) on the mediator (*M=LoAtt*), and this specification is the same as our baseline regression specified in Eq. (2). Thus, the coefficient $\boldsymbol{\alpha_{1,i}}$ captures the effect of *Digit* on *LoAtt*, as reported earlier in Table 3. Eq. (5C) investigates the effects of the independent variable (*X=Digit*) and mediator (*M=LoAtt*) on the dependent variable (*Y=Att-Co*). Thus, the coefficient $c'_{1,i}$ in Eq. (5) captures the direct effect of *Digit* on *Att-Co*, and the product of $\alpha_{1,i}\times b_{1,i}$, from Eq. (5B) and Eq. (5C), measures the indirect effect of *Digit* on attention comovement through *LoAtt* as a mediator.

Columns 1, 2, and 3 of Table 8 report the regression results of Eq. (5A). The coefficients on *Digit* are 0.170, 0.169, and 0.115 in columns 1, 2, and 3, respectively, and these numbers are statistically significant at the 0.1 or higher levels. This evidence indicates that DIT development positively affects attention comovement and stock return correlation. In terms of economic significance, an increase in *Digit* by one standard deviation (0.207) is associated with an increase in attention comovement by approximately 11.0% (=0.170×0.207/0.320) based on the absolute mean value of *Att-Co-SV* (0.320) or an increase in stock correlation of 6.5% (=0.115×0.207/0.367) based on the absolute mean value (0.367) of *Return-Co*.

Columns 4, 5, and 6 of Table 8 report the regression results of Eq. (5C). The coefficients on *Digit* remain positive (0.174, 0.185, and 0.11) and are statistically significant at 0.1 or higher levels. Additionally, the coefficients on *LoAtt* are also positive (0.083, 0.074, and 0.023) but statistically at the 0.05 level in Columns 4 and 5 and statistically insignificant at the 0.1 level in Column 6. This evidence indicates that both *Digit* and local attention bias affect attention comovement and stock return correlation.

For simplicity, we use only the results reported in Column (4) as an example to explain the results associated with DIT's direct and indirect impact. The coefficient on *Digit* (0.174) means that an increase in *Digit* by one standard deviation (0.207) is associated with a direct increase in attention comovement by approximately 11.3% (=0.174×0.207/0.320) based on the absolute mean value (0.320) of *Att-Co-SV*. The coefficient on *LoAtt* (0.083) means that an increase in *LoAtt* by one standard deviation (0.503) is associated with an increase in attention comovement by approximately 13.0% (=0.083×0.503/0.320) based on the absolute mean value of *Att-Co-SV* (0.320). The indirect effect of *Digit* on *Att-Co-SV* is computed as the product of 0.158×0.083, where 0.158 is $\alpha_{1,i}$, the coefficient on *Digit* in Eq. (5B) reported in column (3) of Table 3. In terms of economic significance, an increase in *Digit* by one standard deviation (0.207) is associated with an indirect increase in attention comovement by approximately 0.85% (=0.158×0.083×0.207/ 0.320) based on the absolute mean value of *Att-Co-SV* (0.320), which is caused by *LoAtt* as a mediator. In relative terms, the indirect effect of *Digit* on attention comovement through local attention bias is approximately 7.5% (=0.158×0.083/0.174) of the direct effect.

6. Discussion and conclusions

Although an emerging literature shows that investors pay more attention to and acquire more information about firms in their local area than about nonlocal firms, which is due to investors' information endowment advantage, there is limited evidence on whether the advancement in digital and information technology attenuates or amplifies the local attention bias as information becomes readily available as DIT develops. Investigating this issue is important because we are living in a digital era and investor information acquisition directly affects not only capital allocation but also equity market efficacy and price discovery, especially given that theories offer contradicting predictions regarding investor information acquisitions. We use publicly listed A-share companies in China and observe the following main results.

First, our results show that DIT development proxied by *Digit* positively affects local attention bias. The results are consistent and remain robust based on alternative measures of local attention bias and after controlling for various firm characteristics and local geographic factors. We also use two different interaction IV variables for *Digit* and obtain consistent results, suggesting that the positive effect of DIT on local bias is unlikely due to omitted variables. Second, we discover geographic heterogeneity effects. Our results show that economic development and institutional environment amplify the local attention bias because the positive effect of *Digit* on local attention bias exists only for firms located in provinces with higher economic development measured by GDP per capita and in provinces with a better developed institutional environment. Third, we find that local attention bias positively influences attention comovement and stock return correlation, and the effects are significant both statistically and economically. Fourth, our mediating analysis shows that *Digit* affects attention comovement and stock return correlation both directly and indirectly through local attention bias as a mediator.

Two competing theories offer opposite predictions in investor information acquisition. Our novel findings support strategic complementary theory. Additionally, our empirical findings are consistent with Van Nieuwerburgh and Veldkamp's (2009) theoretical prediction that information immobility persists not because investors cannot learn what locals know but because investors do not choose to learn other others know. Moreover, our results imply that digital and information technology possibly intensifies investors' selective attention and exposure and makes investors further exposed to more polarized information. We also believe that using samples from other markets may generate more fruitful results because investors behave differently in different market environments and cultural influences.

References:

Agarwal, A., A.C. Leung, P. Konana, and A. Kumar, 2017. Cosearch attention and stock return predictability in supply chains. *Information Systems Research* 28(2), 265-288.

Andrei, D., and M. Hasler, 2020. Dynamic attention behavior under return predictability. *Management Science*, 66 (7), 2906-2928.

Banerjee, S., J. Davis, and D. Gondhi, 2018. When transparency improves, must prices reflect fundamentals better? Review of Financial Studies, 31(6), 2377–2414.

Barberis, N., A. Shleifer, and J. Wurgler, 2005. Comovement. *Journal of Financial Economics* 75 (2), 283-317.

Barlevy, G., and P. Veronesi, 2000. Information acquisition in financial markets. *Review of Economic Studies* 67, 79-90.

Baron, R., and D. Kenny, 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.

Birch K (2020) Automated neoliberalism? The digital organization of markets in technoscientific capitalism. *New Formations* 100(100–101), 10–27

Bond, P., A. Edmans, I. Goldstein, 2012. The real effects of financial markets. *Annual Review of Financial Economics* 4, 339–360.

Cheng, X., and J. Lu, 2019. Do local investors have information advantage? An empirical study with Baidu search. *Chinese Journal of Management Science* 27 (4), 25-36.

Cookson, J.A., J.E. Engelberg, W. Mullins, 2023. Echo chambers. *The Review of Financial Studies* 36, 450-500.

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

Cziraki, P., J. Mondria, and T. Wu, 2021. Asymmetric attention and stock returns. *Management Science* 67 (1), 48-71.

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

DeMoya, J.F, and J. Pallud, 2020. From panopticon to heautopticon: A new form of surveillance introduced by quantified self practices. Information Systems Journal 30(6):940–976.

Drake, M., J. Jennings, D. Roulstone, J. Thornock, 2017. The comovement of investor attention. *Management Science* 63 (9), 2847-2867.

Dugast, J., and T. Foucault, 2018. Data abundance and asset price informativeness. *Journal of Financial of Economics* 130(2), 367–391.

Ehrmann, M., and D. Jansen, 2022. Stock return comovement when investors are distracted: more and more homogenous. *Journal of International Money and Finance* 129, 102742.

Feng, J., J. Goodell, M. Li, and Y. Wang, 2023. Environmental information transparency and green innovations. *Journal of International Financial Markets, Institutions & Money* 86, 101799.

Festinger, L. 1957. A theory of cognitive dissonance. Evanston, IL: Row, Peterson & Company.

Goldstein, I., and L. Yang, 2015. Information diversity and complementarities in trading and information acquisition. *Journal of Finance* 70 (4), 1723-1765.

Golman, R., D. Hagmann, and G. Loewenstein, 2017. Information avoidance. *Journal of Economic Literature* 55 (1), 96-135.

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

Guo, F., J. Wang, F. Wang, Z. Cheng, T. Kong, and X. Zhang, 2020.Measuring China's digital financial inclusion: index compilation and spatial characteristics. *China Economic Quarterly* 19 (4), 1401-1418.

Hellwig, C., and L. Veldkamp, 2009. Knowing what others know: coordination motives in information acquisition. *Review of Economic Studies* 76, 223-251.

Hirshleifer, D. 2015. Behavioral finance. *Annual Review of Financial Economics* 7, 133-159.

Hirst, D.E., and P.E. Hopkins, 1998. Comprehensive income reporting and analysts' valuation judgment. *Journal of Accounting Research* 36, 47-75.

Huang, S., Y. Huang, and T. Lin, 2019. Attention allocation and return comovement: evidence from repeated natural experiments. *Journal of Financial Economics* 132, 369-383.

Huang, Y., H. Qiu, and Z. Wu, 2016. Local bias in investor attention: evidence from China's internet stock message boards. *Journal of Empirical Finance* 38, 338-354.

Huang, S., and B. Yueshen, 2021. Speed acquisition. *Management Science* 67 (6), 3492-3518.

Ivkovic, Z., and S. Weisbenner, 2005. Local does as local is: information content of the geography of individual investors' common stock investments. *Journal of Finance* 60(1), 267–306.

Jiang, Z., A. Rai, H. Sung, C. Nie, and Y. Hu, 2023. How does online information influence offline transactions? Insights from digital real estate platforms. *Information System Research*, https://doi.org/10.1287/isre.2020.0658

Karlsson, N., G. Loewenstein, and D. Seppi, 2009. The ostrich effect: selective attention to information. *Journal of Risk and Uncertainty* 38, 95-115.

Kendall, C., 2018. The time cost of information in financial markets. *Journal of Economic Theory* 176, 118-157.

Knobloch-Westerwick, S. 2014. Choice and preference in media use: Advances in selective exposure theory and research. England, UK: Routledge.

Levy, R., 2021. Social media, news consumption, and polarization: evidence from a field experiment. *American Economic Review* 111 (3), 831-870.

Lewis, K., 1999. Trying to explain home bias in equities and consumption. *Journal of Economic Literature* 37, 571-608.

Li, M., D. Liu, H. Peng, and L. Zhang, 2020. Does low synchronicity mean more or less informative prices? Evidence from an emerging market. *Journal of Financial Stability* 51, 100817.

Mendelson H, and R.R. Pillai, 1998. Clock speed and informational response: Evidence from the information technology industry. *Information Systems Research* 9(4), 415–433.

Morck, R., B. Yeung, and W. Yu, 2000. The information content of stock markets: why do emerging markets have synchronous price movements? *Journal of Financial Economics* 58(1-2), 215-260.

Ngwenyama, O., F. Rowe, S. Klein, H. Henriksen, 2023. The Open Prison of the Big Data Revolution: False Consciousness, Faustian Bargains, and Digital Entrapment. *Information Systems Research*, https://doi.org/10.1287/isre.2020.0588

Nunn, N., and N. Qian, 2014. U.S. food aid and civil conflict. *American Economic Review* 106 (6), 1630-1666.

Peng, L., 2005. Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis* 40 (2), 307-329.

Peress, J., 2004. Wealth, information acquisition, and portfolio choice. *Review of Financial Studies* 17 (3), 879-914.

Scherer, A. M., P.D. Windschitl, and A.R. Smith. 2013. Hope to be right: biased information seeking following arbitrary and informed predictions. *Journal of Experimental Social Psychology* 49 (1), 106–12.

Shiffrin, R., and W. Schneider, 1977. Controlled and automatic human information process: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review* 84 (2), 127-190.

Sicherman, Nachum, George Loewenstein, Duane J. Seppi, and Stephen P. Utkus. 2016. Financial Attention. *Review of Financial Studies* 29 (4): 863–97.

Sims, C., 2003. Implications of rational inattention. *Journal of Monetary Economics* 50, 665-690.

Van Nieuwerburgh, S., and L. Veldkamp, 2009. Information immobility and the home bias puzzle. *Journal of Finance 56 (5), 678-695*.

Van Nieuwerburgh, S., and L. Veldkamp, 2010. Information acquisition and under diversification. *Review of Economic Studies* 77, 779-805.

Veldkamp, L., 2006a. Information markets and the comovement of asset prices. *Review of Economic Studies* 73, 823-845.

Veldkamp, L., 2006b. Media frenzies in markets for financial information. *The American Economic Review* 96 (3), 577-601.

Verrecchia, R.E., 1982. Information acquisition in a noisy rational expectations economy. *Econometrica 50,* 1415–1430 .

Wang, X. L., G. Fan, and L. P. Hu, 2018. Marketization Index of China's Province: NERI Report. Beijing, China: Social Science Academics Press.

Wen, Z., and B. Ye, 2014. Analysis of mediating effects: the development of methods and models. *Advances in Psychological Science* 22 (5), 731-745.

Zhao, T., Z. Zhang, and S. Liang, 2020. Digital economy, entrepreneurship, and high-quality economic development: empirical evidence from urban China. *The Management World* (Chinese) 10, 65-76.

Table 1: Definition of main variables

Variable	*Definition*
LoAtt	Local attention computed as $Ln\ [1+ (S_{i,j,t}/S_{i,t})/(IAP_{j,t}/IAP_t)]$, where $S_{i,j,t}$ is the number of internet (Baidu) searches for *firm i* headquartered in province *j* by people in province *j* in year *t*. $S_{i,t}$ is the number of internet searches for *firm i* by people in all provinces in year *t*. $IAP_{j,t}$ is the number of internet access ports in province *j* in which firm *i* is headquartered in year *t*, and IAP_t is the number of internet access ports in all provinces in year *t* (Huang et al., 2016; Cheng and Lu, 2019).
LoAttx	An alternative measure of local attention, computed as $Ln\ [1+(S_{i,j,t}/S_{i,t})/(IU_{j,t}/IU_t)]$. $S_{i,j,t}$ and $S_{j,t}$ are the same as defined above. $IU_{j,t}$ and IU_t are the number of internet users in province *j* in which firm *i* is headquartered and in all provinces, respectively, in year *t* (Huang et al., 2016).
Digit	Digitalization index in province *j* in which firm *i* is headquartered in year *t*. It is computed as the principal component of five categories of data: the number of internet users per 100 people; percentage of labor force specializing in internet and computer service sectors; industry output of telecommunication and information technology; percentage of people using mobile phones; digital financial inclusion index developed by the Research Center of Digital Finance of Beijing University (Guo et al., 2020; Zhao et al., 2020).
Att-Co-FA	Attention comovement based on the number of forecast updates by financial analyst who cover firm *i* in year *t* (Drake et al., 2017).
Att-Co-SV	Attention comovement based on Baidu search volume for firm in year t (Drake et al., 2017).
Return-Co	Stock return comovement based on weekly returns (Drake et al., 2017; Morck et al., 2000).
Size	Ln(1+total assets) measuring firm size at the end of year *t*.
ROA	Return on total assets of firm *i* in year *t*.
Leverage	Leverage ratio (debts/total assets) of firm *i* in year *t*.
BK/Mkt	The book value of equity divided by the market value of equity of firm *i* in year *t*.
Return	Annual stock return (including dividend) of firm *i* in year *t*.
SaleGrowth	Sales growth rate of firm *i* in year *t*, computed as $Sales_t/Sales_{t-1} – 1$.
InstSH(%)	Percentage of firm *i*'s shares owned by all institutions in year *t*.
SOE	A dummy variable taking a value of 1 if the controlling shareholder is either a local or state-government in year *t* and 0 otherwise.
Analyst	Logarithm of 1 plus the number of analyst report for firm *i* in year *t*.
HS300-Index	Equals 1 if firm *i* is included in HS00-index in year *t* and 0 otherwise.
Ln(TrdVol)	Logarithm of the number of shares trading for firm *i* in year *t*.
GDP/P	GDP per capita in province *j* in which firm *i* is headquartered in year *t*.
CollegeP(%)	College students as a percentage of total population in province *j* in which firm *i* is headquartered in year *t*.
PopGrowth(%)	Population growth rate in province *j* in which firm *i* is headquartered in year *t*.
CenterCity	A dummy variable taking a value of 1 if firm I is headquartered in one of the following center cities in year t and 0 otherwise. The center cities include Beijing, Tianjin, Shanghai, Guangzhou, Chongqing, Chengdu, Wuhan, Zhengzhou, and Xian.

Table 2: Description of main variables

	N	Mean	StdDev	Min	p25	p50	p75	Max
LoAtt	10450	1.972	0.503	0.000	1.622	1.906	2.264	4.349
LoAttx	6314	1.837	0.458	0.493	1.506	1.811	2.122	4.032
Digit	10450	0.197	0.207	0.004	0.047	0.115	0.273	0.819
Att-Co-FA	10546	-2.132	1.465	-6.602	-2.957	-1.829	-1.046	0.343
Att-Co-SV	10546	-0.320	1.008	-3.656	-0.867	-0.176	0.377	1.524
Return-Co	10546	-0.367	0.917	-3.291	-0.916	-0.284	0.276	1.461
Size	10383	22.890	1.360	16.520	21.931	22.709	23.656	28.636
Leverage	10383	0.473	0.218	0.008	0.322	0.475	0.619	8.256
ROA	10546	0.043	0.064	-1.495	0.017	0.038	0.068	0.590
SaleGrowth	10546	1.197	0.510	0.325	0.996	1.111	1.262	7.422
BK/Mkt	10546	0.526	0.378	-0.044	0.261	0.425	0.681	2.466
Return	10546	0.147	0.576	-0.766	-0.215	0.015	0.357	15.211
SOE	10375	0.485	0.500	0.000	0.000	0.000	1.000	1.000
Analyst	10546	24.205	28.766	1.000	5.000	13.000	33.000	290.00
InstSH(%)	10546	47.816	22.728	0.007	31.782	49.869	65.067	326.73
HS300_Index	10546	8.370	24.475	-25.310	-12.090	5.580	27.210	51.660
Ln(TrdVol)	10546	21.577	0.993	18.226	20.910	21.542	22.220	26.138
GDP/P	10383	71.148	33.091	16.024	45.723	64.516	88.521	164.16
CollegeP(%)	10383	0.005	0.001	0.002	0.004	0.005	0.006	0.010
PopGrowth(%)	10078	4.910	2.504	-4.480	2.970	4.940	6.560	11.470
CenterCity	10546	0.292	0.455	0.000	0.000	0.000	1.000	1.000

This table reports the summary statistics of the main variables at the firm-year level during the sample period of 2011-2020 except for *LoAttx*, which is during the period of 2012-2016 due to limited data. All variables are defined in Table 1.

Table 3: Regression of digitalization on local attention

	(1)	(2)	(3)	(4)
Digit	0.181***	0.173***	0.160***	0.158***
	(0.031)	(0.031)	(0.030)	(0.030)
Size		-0.058***	-0.011**	-0.012**
		(0.004)	(0.005)	(0.005)
Leverage		0.073**	0.030	0.030
		(0.029)	(0.025)	(0.025)
ROA		0.264***	-0.109*	-0.067
		(0.063)	(0.058)	(0.056)
SaleGrowth		0.010	-0.007	-0.006
		(0.006)	(0.006)	(0.006)
BK/Mkt		0.075***	0.065***	0.063***
		(0.012)	(0.012)	(0.012)
Return		-0.026***	-0.018***	-0.014***
		(0.006)	(0.005)	(0.005)
SOE		0.028***	0.028***	0.028***
		(0.007)	(0.007)	(0.007)
Analyst			0.001***	0.001***
			(0.000)	(0.000)
InstSH(%)			0.000*	0.000**
			(0.000)	(0.000)
HS300_Index			0.016***	0.004***
			(0.000)	(0.001)
Ln(TrdVol)			-0.135***	-0.134***
			(0.004)	(0.004)
GDP/P				0.012***
				(0.001)
CollegeP(%)				30.649*
				(18.278)
PopGrowth(%)				-0.021***
				(0.003)
CenterCity				0.015
				(0.010)
Constant	1.983***	3.148***	5.370***	3.851***
	(0.025)	(0.075)	(0.088)	(0.216)
Ind./Year/Location FE	Yes	Yes	Yes	Yes
N	10450	10287	10287	9917
R^2-Adj.	0.680	0.696	0.737	0.753

This table reports the ordinary least squares (OLS) regression of digitalization (*Digit*) on local attention measured by *LoAtt*. All continuous variables are winsorized at 1% and 99% to reduce data noise. Standard errors in parentheses, *, **, *** indicate 0.1, 0.05, and 0.01 levels of significance, respectively.

Table 4: 2SLS regression with different instrumental variables

	$IV=Ph_{84}\times IU_{t-1}$		$IV=RDLS\times IU_{t-1}$	
	1st stage	2nd state	1st stage	2nd state
	Digit	*LoAtt*	*Digit*	*LoAtt*
	(1)	*(2)*	*(3)*	*(4)*
$IV(=Ph_{84}\times IU_{t-1})$	0.021***			
	(0.000)			
Digit (IVed1)		1.573***		
		(0.053)		
$IV(=RDLS\times IU_{t-1})$			0.008***	
			(0.000)	
Digit (IVed2)				1.479***
				(0.192)
Constant	-0.068	5.179***	-0.688***	4.998***
	(0.053)	(0.170)	(0.059)	(0.201)
Control variables	Yes	Yes	Yes	Yes
Ind./Year/Location FE	Yes	Yes	Yes	Yes
N	9718	9718	9908	9908
R^2-Adj.	0.557	0.232	0.393	0.239
Kleibergen–Paap rk LM stat		2493.59		461.38
Stock-Yogo weak ID test		3968.92		886.69
10% maximal IV size		16.38		16.38

This table reports 2SLS regressions with two different instrumental variables. The first IV is an interaction variable of $Ph_{84}\times IU_{t-1}$ (Column 1), where Ph_{84} is the number of landline phone users per ten thousand persons in a province in 1984, a time-invariant proxy of the province's historical telecommunication technology. IU_{t-1} is the number of internet users in the whole country in year t-1, a time-variant proxy for overall telecommunication development. The second IV is an interaction variable of $RDLS\times IU_{t-1}$ I (Column 3), where *RDLS* is the relief degree of the land surface in a province, an alternative time-invariant proxy for the province's telecommunication technology. *Digit(IVed1)* and *Digit(IVed2)* are instrumented *Digit* from columns 1 and 3, respectively. The control variables are the same as those reported in Table 3. All continuous variables are winsorized at 1% and 99% to reduce data noise. Standard errors in parentheses, *, **, *** indicate 0.1, 0.05, and 0.01 levels of significance, respectively.

Table 5: Regression of digitalization using an alternative measure of local attention bias

	(1)	(2)	(3)	(4)
Digit	0.167***	0.143***	0.153***	0.136***
	(0.044)	(0.043)	(0.040)	(0.040)
Size		-0.060***	-0.007	-0.007
		(0.005)	(0.005)	(0.005)
Leverage		0.076**	0.009	0.007
		(0.035)	(0.024)	(0.023)
ROA		0.513***	-0.012	-0.011
		(0.082)	(0.062)	(0.061)
SaleGrowth		0.011*	-0.010	-0.013**
		(0.007)	(0.006)	(0.006)
BK/Mkt		0.076***	0.066***	0.071***
		(0.015)	(0.014)	(0.014)
Return		-0.017***	-0.013***	-0.010**
		(0.006)	(0.005)	(0.005)
SOE		0.028***	0.032***	0.033***
		(0.007)	(0.007)	(0.007)
Analyst			0.001***	0.001***
			(0.000)	(0.000)
InstSH(%)			0.001***	0.001***
			(0.000)	(0.000)
HS300_Index			0.049***	0.027***
			(0.001)	(0.004)
Ln(TrdVol)			-0.155***	-0.156***
			(0.005)	(0.005)
GDP/P				0.012***
				(0.001)
CollegeP(%)				-26.899
				(19.270)
PopGrowth(%)				-0.005
				(0.004)
CenterCity				0.006
				(0.009)
Constant	2.005***	3.216***	6.472***	5.021***
	(0.026)	(0.091)	(0.106)	(0.322)
Ind./Year/Location FE	Yes	Yes	Yes	Yes
N	6314	6209	6209	6209
R^2-Adj.	0.713	0.735	0.795	0.804

This table reports the ordinary least squares (OLS) regression of digitalization (*Digit*) on local attention measured by *LoAtt*x, an alternative measure of local attention (Huang et al., 2016). See Table 1 for detailed information. The sample period is 2011-2016 because data for computing *LoAttx* are not available after 2016. All continuous variables are winsorized at 1% and 99% to reduce data noise. Standard errors in parentheses, *, **, *** indicate 0.1, 0.05, and 0.01 levels of significance, respectively.

Table 6: Regression of digitalization on local attention considering heterogeneity

	(1)	(2)	(3)	(4)
Panel A: Subsamples based on economic development				
	Undeveloped	Undeveloped	Developed	Developed
Digit	0.008	-0.065	0.291^{***}	0.225^{***}
	(0.079)	(0.084)	(0.033)	(0.035)
Constant	1.782^{***}	4.632^{***}	1.942^{***}	4.575^{***}
	(0.056)	(0.287)	(0.025)	(0.205)
Controls variables	No	Yes	No	Yes
Ind./Year/Location FE	Yes	Yes	Yes	Yes
N	2495	2407	7955	7510
R^2-Adj.	0.494	0.551	0.761	0.836
Panel B: Subsamples based on institutional environment – marketization index (MI)				
	MI-Low	MI-Low	MI-High	MI-High
Digit	0.023	0.085^{*}	0.342^{***}	0.275^{***}
	(0.047)	(0.046)	(0.042)	(0.046)
Constant	1.749^{***}	4.802^{***}	2.000^{***}	2.416^{***}
	(0.037)	(0.174)	(0.040)	(0.281)
Controls variables	No	Yes	No	Yes
Ind./Year/Location FE	Yes	Yes	Yes	Yes
N	5503	5413	4947	4504
R^2-Adj.	0.556	0.614	0.816	0.889
Panel C: Subsamples based on digitalization level				
	Digit-Low	Digit-Low	Digit-High	Digit-High
Digit	0.090	0.069	0.212^{***}	0.252^{***}
	(0.056)	(0.045)	(0.059)	(0.061)
Constant	1.803^{***}	2.797^{***}	2.615^{***}	5.403^{***}
	(0.031)	(0.312)	(0.036)	(0.375)
Controls variables	No	Yes	No	Yes
Ind./Year/Location FE	Yes	Yes	Yes	Yes
N	4994	4918	5456	4999
R^2-Adj.	0.703	0.799	0.653	0.703

This table reports the ordinary least squares (OLS) regression of digitalization (*Digit*) on local attention measured by *LoAtt* under different situations. In Panel A, firms are classified as Undeveloped (Developed) if the GPD per capita of the province in which the firm is headquartered is below (equal to or above) the mean value of GDP per capita of all provinces in year *t*. In Panel B, firms are classified as MI-Low (MI-High) if the marketization index (MI) of the province in which the firm is headquartered is below (equal to or above) the mean value of MI of all provinces in year *t*. In Panel C, observations before 2016 (including 2016) are classified as Digit-Low since there is a steady increase in Digit after 2016, and observations after 2016 are classified as Digit-High. The control variables are the same as those reported in Table 3. All continuous variables are winsorized at 1% and 99% to reduce data noise. Standard errors in parentheses, *, **, *** indicate 0.1, 0.05, and 0.01 levels of significance, respectively.

Table 7: Regression of local attention bias on attention comovement and return comovement

	(1)	(2)	(3)	(4)	(5)	(6)
	Att-Co-SV	*Att-Co-SV*	*Att-Co-FA*	*Att-Co-FA*	*Return-Co*	*Return-Co*
LoAtt	0.103***	0.131***	0.089*	0.099*	0.178***	0.129***
	(0.033)	(0.038)	(0.051)	(0.057)	(0.029)	(0.031)
Constant	-0.625***	-3.235***	-2.297***	-5.390***	-0.588***	2.111***
	(0.118)	(0.624)	(0.178)	(0.902)	(0.091)	(0.526)
Control variables	*No*	Yes	*No*	Yes	*No*	Yes
Ind./Year/Location FE	Yes	Yes	Yes	Yes	Yes	Yes
N	10450	9917	10450	9917	10450	9917
R^2-Adj.	0.1144	0.2099	0.0298	0.1461	0.2012	0.3198

This table reports regressions of local attention on attention comovements and return comovements. In columns 1 and 2, the dependent variable is *Att-Co-SV*, measuring attention comovement computed based on internet (Baidu) search volume. In columns 3 and 4, the dependent variable is *Att-Co-FA*, measuring attention comovement computed based on the forecast updates of financial analysts who cover the firm (Drake et al., 2017). In columns 5 and 6, the dependent variable is *Return-CO*, measuring return comovement computed based on weekly returns of individual stocks, market index, and industry (Morck et al., 2000). The control variables are the same as those reported in Table 3. All continuous variables are winsorized at 1% and 99% to reduce data noise. Standard errors in parentheses, *, **, *** indicate 0.1, 0.05, and 0.01 levels of significance, respectively.

Table 8: Mediation analysis of local attention bias

	(1)	(2)	(3)	(4)	（5）	（6）
	Att-Co-SV	*Att-Co-FA*	*Return-Co*	*Att-Co-SV*	*Att-Co-FA*	*Return-Co*
Digit	0.170^{**}	0.169^{*}	0.115^{**}	0.174^{**}	0.185^{**}	0.110^{*}
	(0.068)	(0.093)	(0.058)	(0.068)	(0.094)	(0.058)
LoAtt				0.083^{***}	0.074^{**}	0.023
				(0.026)	(0.037)	(0.023)
Constant	-2.215^{***}	-4.465^{***}	0.811^{***}	-2.565^{***}	-4.231^{***}	0.737^{***}
	(0.318)	(0.439)	(0.258)	(0.341)	(0.455)	(0.272)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Ind./Year/Location FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9917	9917	9917	9917	9917	9917
R^2-Adj.	0.2135	0.1467	0.1930	0.2142	0.1469	0.1931

This table reports a mediation analysis of local attention bias on attention comovements and return comovements. Columns 1, 2, and 3 report the regression results of Eq. (5A), and columns 4-6 report the results of Eq. (5C). The dependent variables are Att-Co-SV, Att-Co-FA, and Return-Co, which are attention comovement measured by Baid search volume, attention comovement measured by volume financial analyst forecast updates, and stock return correlation. The control variables are the same as those reported in Table 3. All continuous variables are winsorized at 1% and 99% to reduce data noise. Standard errors in parentheses, *, **, *** indicate 0.1, 0.05, and 0.01 levels of significance, respectively.