Information externalities, analyst research resource allocation, and stock pricing efficiency

Chenyu Cui\(^a\), Guihua He\(^b\) and Deren Xie\(^c\)

\(^a\)School of Business, University of International Business and Economics, Beijing, China; \(^b\)School of Accounting, Zhongnan University of Economics and Law, Wuhan, China; \(^c\)School of Economics and Management, Tsinghua University, Beijing, China

**ABSTRACT**
Information is key to decision-making and is a major determinant of investment performance. We hypothesise that as analysts are constrained by their research resources, they collect information with more externalities. Measuring information externalities as a stock’s fundamental correlations with other stocks in the same industry, we find that stocks with high information externalities get more analyst coverage, more (high-quality) analyst reports, and more site visits than those with low information externalities. At the analyst level, we find consistent evidence that individual analysts allocate more effort to firms with high information externalities in their stock portfolios. Consequently, we further demonstrate that stocks with high information externalities are associated with less information delay and are priced more efficiently.

**KEYWORDS**
Information externalities; research resource allocation; pricing efficiency; analyst

1. Introduction

The efficient market hypothesis (Fama, 1970) states that information can always be injected into stock prices in a timely, sufficient, and accurate manner. However, in practice, investors spend a lot of time and energy searching and processing information; therefore, stock prices can be only partially efficient under the constraint of information acquisition costs (Grossman & Stiglitz, 1980). Thousands of stocks are listed and traded in today’s stock markets; thus, it is important to understand how investors allocate scarce resources to research and acquire valuable information. In this study, we examine the effect of information externalities on the information production of sell-side analysts.

Economic theories point out that information is a kind of special goods because it has a high acquisition cost while its replication cost is negligible (Romer, 1990; Veldkamp, 2006). Therefore, the value of information hinges on whether that information is useful for a set of stocks or just one stock. Clearly, the larger the set of stocks, the higher the value of information. In the context of stock trading, the value of the information obtained from one particular stock lies in two dimensions. The first refers to the notion that information collected from a stock always reflects its fundamentals. The second dimension is that...
information can be used to price other stocks, for example, other stocks in the same industry, geographic region, or supply chain. In this study, we consider the value of information reflected in the second dimension as information externalities. As the replication cost of information is negligible compared with its acquisition cost, if that information can be applied to a sufficiently large number of stocks, then information externalities will be sufficiently high. Therefore, we assume that analysts will allocate more research resources to stocks with high information externalities.

Following the design of Hameed et al. (2015), we measure a firm’s information externalities as the correlation between its fundamentals (return on assets or ROA) and the fundamentals of other firms within the same industry. If a firm has strong fundamental correlations with its industry peers, by devoting research resources to that firm, analysts can obtain information that is valuable not only for the firm itself but also for other firms in the same industry. However, if analysts allocate their research effort to firms with weak fundamental correlations, the information acquired from that firm can only be used to price just a few stocks.

Although analysts are not the only market participants who are concerned about the external value of information, they are the most suitable subjects for us to examine the relationship between information externalities and the allocation of limited research resources for the following reasons. First, as proposed in prior literature (Driskill et al., 2020; Harford et al., 2019), due to constraints on attention and cognitive capacity, analysts need to optimise their resource allocation to maximise their utility function. Second, analysts are an important information intermediary in stock markets. They provide research and investment recommendations to meet the information needs of investors, and their work has a substantial impact on stock prices. Third, the availability of analyst data allows us to quantify analysts’ resource allocation in multiple aspects, such as stock coverage and the frequency, quality, and timeliness of forecasts. Furthermore, analysts make forecasts for multiple firms, which allows us to explore how individual analysts balance their research resources within their portfolios.

Using a large sample of A-share listed firms from 2007 to 2018, we find that firms with higher information externalities are covered by more analysts. Analysts also issue more reports, revise earnings forecasts more frequently, and conduct more site visits for these firms. We also explore research resource allocation at the individual analyst level. By adding analyst × year fixed effects to analyst-level regressions, we can estimate whether analysts are selective in their resource allocation using variations within their portfolios. We find that analysts produce more in-depth reports and revise earnings forecasts more frequently for firms with high information externalities than for those with low information externalities. We also find that analyst forecast revisions are timelier and more informative for such firms. Collectively, these findings suggest that analysts expend their research resources strategically and devote more effort to researching firms with higher information externalities.

It is well known that pricing efficiency is largely determined by investors’ information acquisition incentives (Grossman & Stiglitz, 1980; Verrecchia, 1982), and the information production of sell-side analysts strongly contributes to firms’ information environment (Derrien & Kecskés, 2013; Harford et al., 2019). Therefore, we further examine whether stocks with different levels of information externality have different pricing efficiency. Given that firms with higher information externalities receive more research resources
from analysts, we expect industry information to be integrated into the price of stocks with high information externalities rapidly and with a significant delay for stocks with low information externalities. The above prediction is confirmed by the lead-lag effect in time series tests. We find that the lagged returns of stocks with high information externalities significantly predict the current returns of stocks with low information externalities, but the lagged returns of stocks with low information externalities have no explanatory power for the current returns of high information externality stocks. In addition, we find consistent evidence in cross-sectional multivariate regression tests, in which we regress the firm-specific information delay measures developed by Hou and Moskowitz (2005) on information externalities and find significant and negative coefficients.

Next, we use earnings announcements to examine whether investors pay more attention to stocks with high information externalities. If so, we expect the earnings information of high information externality stocks to also be useful in pricing the stocks of their industry peers. Consistent with our hypothesis, we find that the earnings announcements of high information externality stocks have greater spillover effects on other stocks within the same industry than those of low information externality stocks.

We find that information externalities also exhibit within geographic regions, and supply chains. Using a province information externality measure, we replicate the main empirical tests and obtain similar findings. Furthermore, due to the close economic links within supply chains, we hypothesise that a stock’s information externalities will increase significantly after the initial public offering (IPO) of any of its main customers or suppliers. Taking the IPO of a firm’s customers or suppliers as exogenous shocks and using a difference-in-differences approach, we find that firms receive significantly more research resources from analysts after the IPO of its customers or suppliers.

Finally, the importance of information externalities should be recognised by all stock market participants who trade a large number of stocks, not just by analysts. Although the behaviour of stock market participants other than analysts is usually unobservable, we manage to test it by using site visit data. We find that mutual funds and other institutional investors are more likely to conduct site visits to firms with high information externalities, which further supports the validity of our information externality theory for other market participants.

Our study makes several contributions to the literature. First, we investigate the research resource allocation of sell-side analysts in China’s A-share market and present information externality as a new determinant of such allocation. In this way, our study extends the theoretical and empirical findings of Peng and Xiong (2006), Kacperczyk et al. (2014), Kacperczyk et al. (2016), and Harford et al. (2019).

Second, our study contributes to the literature on limited attention and rational inattention. It is generally accepted that investors with limited attention are prone to ignore valuable information. However, the literature ignores the possibility of investors strategically allocating their limited attention by choosing to be rationally inattentive to some stocks. Our study, together with those of Peng and Xiong (2006) and Van Nieuwerburgh and Veldcamp (2009), demonstrates that many market-related topics, such as portfolio choice, stock price synchronicity, and pricing efficiency, can be the endogenous result of investors’ optimal resource allocation strategy under time, attention, and energy constraints.
Third, prior literature shows that pricing efficiency depends on the information acquisition incentives of investors, which in turn depends on information acquisition costs (Grossman & Stiglitz, 1980; Verrecchia, 1982). We add to this literature by demonstrating that information externalities are also an important determinant of information acquisition incentives, and thus advance our understanding of market efficiency.

The remainder of this paper is organised as follows. Section 2 reviews the related literature and develops the hypotheses. Section 3 describes the research design. Section 4 presents and discusses the main empirical results. Section 5 provides additional evidence and Section 6 concludes the paper.

2. Literature review and hypothesis development

Since the pioneering study of Kahneman (1973), it has been widely recognised that investors have limited attention; many studies in various settings provide compelling evidence for it. For example, investors’ reaction to earnings announcements is weaker and the post-earnings announcement drift (PEAD) is stronger when investors are overloaded with many same-day earnings announcements made by different firms (Hirshleifer et al., 2009). As investors are more likely to be distracted on Fridays, PEAD is more pronounced for earnings announced on Fridays (DellaVigna & Pollet, 2009).

Although limited attention is often deemed to be a feature of retail investors (Barber & Odean, 2008), a growing literature provides evidence that professional investors are also constrained by it. Kempf et al. (2017) find that institutional investors are unable to act as corporate monitors when their attention is distracted by industries with extreme returns. Ben-Rephael et al. (2017) show that price drifts following earnings announcements and analyst recommendation changes are largely caused by the insufficient attention of institutional investors. Driskill et al. (2020) demonstrate that analysts update their earnings forecasts less frequently when multiple portfolio firms announce earnings on the same day.

The scarcity of research resources highlights the importance of allocation. Peng and Xiong (2006) propose that, due to limited attention, rational investors tend to process more market and industry information than firm-specific information. This type of allocation strategy generates categorical thinking and comovement in stock returns. Huang et al. (2019) use large jackpot lotteries that distract investor attention from the stock market as exogenous shocks, and find that stock price synchronicity is significantly higher on large jackpot days than on other days. Therefore, the findings of Huang et al. (2019) support the theory proposed by Peng and Xiong (2006).

The allocation of research resources is also related to portfolio choice. When choosing stocks to hold, investors are free to acquire information about stocks they are already well informed about (such as local stocks) or learn about unfamiliar stocks. Van Nieuwerburgh and Veldcamp (2009) consider investors’ information acquisition and portfolio decisions jointly and find that investors strategically allocate their research resources to further specialise in what they know well. Their theory effectively explains why investors continue to hold undiversified portfolios tilting towards local stocks even when information is readily available in the modern information era.

Kacperczyk et al. (2014) and Kacperczyk et al. (2016) examine how fund managers allocate research resources at different stages of economic cycles. They find that fund managers devote more research resources to time the macroeconomic trends during
recessions and to stock picking during booms. Research resource allocation is also an important issue for analysts. Clement (1999) points out that an oversized portfolio impedes the quality of analyst earnings forecasts. Harford et al. (2019) examine how the career concerns of analysts affect the way they allocate research resources. They find that analysts allocate more research effort to portfolio stocks that are more important to their careers, for example, stocks that generate more commissions for brokerage houses.

Our study extends the literature by examining how analysts allocate their efforts from a different perspective: information externalities. As pointed out by Romer (1990) and Veldkamp (2006), the most prominent feature of information processing is the scale effect: information has high fixed costs and a near-zero using cost. Therefore, all else being equal, the value of information is mainly determined by its range, which is the number of stocks in a sector (for example, industry) where that information can be used. Analysts mainly provide information services to institutional investors. Although institutional investors value the skill of stock picking, they care much more about the performance of their entire portfolios. Therefore, to better meet the need of their clients, the optimal choice for resource-constrained analysts is to acquire information that can guide the portfolio decisions of institutional investors. In other words, analysts should pay close attention to information externalities.

We visualise the above arguments in Figure 1. Suppose that an analyst needs to advise their client, an institutional investor, and has to decide how to allocate efforts between stock A and stock Z. As shown in the graph, the information acquired from stock A is useful for both stock A and for a set of other stocks including stocks B, C, D, E, and others, then stock A has high information externalities. Conversely, the information collected

![Figure 1. Information externalities and analyst research resource allocation.](image-url)
from stock $Z$ can only be used for its own trading decision, thus stock $Z$ has low information externalities. Hence, the optimal strategy for the analyst is to spend more research effort on stock $A$ to become a better advisor.

Based on the analysis above, we propose the first empirical hypothesis as follows.

**Hypothesis 1:** Ceteris paribus, firms with high information externalities receive more research resources from analysts than firms with low information externalities.

It is well known that stock pricing efficiency is largely influenced by investors’ information acquisition incentives (Grossman & Stiglitz, 1980; Verrecchia, 1982), which in turn is a function of the costs and benefits of acquiring information from that stock. As mentioned above, analysts are likely to exert more effort on stocks with high information externalities; therefore, information externalities are likely to have implications for pricing efficiency.

As information externalities affect firms connected by economic links, such as the same industry or geographic proximity, we consider the diffusion of sector-wide information among stocks of the same sector and pricing efficiency accordingly. Prior literature (Brennan et al., 1993; Chordia & Swaminathan, 2000; Lo & MacKinlay, 1990) shows that some stocks respond to market information rapidly, while others exhibit significant information delay and adjust sluggishly. In other words, the returns of efficiently priced stocks lead the returns of inefficiently priced stocks (lead-lag effect). Hou (2007) finds that, within the same industry, big firms respond to industry information more rapidly than small firms, as big firms have lower information collection costs and more transparent information environments. Similarly, we argue that compared with stocks with low information externalities, stocks with high information externalities are priced more efficiently in the sense of information delay because such firms are associated with higher information collection benefits for analysts and other stock market participants.

Therefore, we propose the second hypothesis as follows.

**Hypothesis 2:** Ceteris paribus, firms with high information externalities are associated with higher pricing efficiency than firms with low information externalities.

3. Research design

3.1. Sample and data

Our initial sample consists of all non-financial companies listed on the China’s A-share stock market. We require quarterly financial data for the last 5 consecutive years to construct $\text{InfoExtVal}$, our information externality measure. As A-share firms started disclosing quarterly financial statements from 2002 onwards, the sample period is from 2007 to 2018, with the firms in the sample having to be listed for no less than 5 years. Financial and site visit data are obtained from the China Stock Market and Accounting Research (CSMAR) database, while analyst data are retrieved from the Suntime database.
3.2. Key variable construction

Following Hameed et al. (2015), we measure InfoExtVal, a stock’s information externalities, as the average correlation between the firm’s fundamentals and the fundamentals of other firms in the same industry. Industries are classified according to the 2001 CSRC (Chinese Securities Regulatory Committee) categories and the manufacturing industry is further divided into the second subcategories.

To calculate InfoExtVal for firm \( k \), we first pair firm \( k \) with firm \( i \) in the same industry \( (k, i) \). We regress firm \( i \)'s ROA\(_{iq} \) on ROA\(_{Mq} \) (the market average ROA excluding that of firm \( i \)) over the previous 5-year window using quarterly financial data in equation (1) to obtain the \( R^2 \) of regression (1), \( R^2_{i,excl,k} \). \( R^2_{i,excl,k} \) measures the extent to which firm \( i \)'s fundamentals can be explained by ROA\(_{Mq} \).

\[
ROA_{iq} = a_i + b_iROA_{Mq} + \epsilon_{iq} \quad (1)
\]

Next, we use equation (2) to regress firm \( i \)'s ROA\(_{iq} \) on ROA\(_{Mq} \) and firm \( k \)'s ROA\(_{kq} \) simultaneously over the same 5-year window to obtain \( R^2_{i,incl,k} \). The \( R^2 \) of regression (2). \( R^2_{i,incl,k} \) represents the incremental explanatory power of firm \( k \) on firm \( i \)'s fundamentals.

\[
ROA_{iq} = a_i + b_iROA_{Mq} + c_iROA_{kq} + \epsilon_{iq} \quad (2)
\]

Then, we extract firm \( k \)'s incremental explanatory fraction on firm \( i \)'s fundamentals using equation (3).

\[
PCORR_{ROA}_{k,i} = \frac{R^2_{i,incl,k} - R^2_{i,excl,k}}{1 - R^2_{i,excl,k}} \quad (3)
\]

Next, we pair firm \( k \) with all other firms in the same industry and rerun equations (1), (2), and (3) for each pair. PCORR\(_{ROA}_{k} \) is the average PCORR\(_{ROA}_{k,i} \) across all \( (k, i) \) pairs. If an industry has \( N \) firms, we will get \( N-1 \) firm pairs matched with firm \( k \).

Finally, InfoExtVal\(_{k} \) is defined as the log-transformed value of PCORR\(_{ROA}_{k} \), as shown in equation (4). A higher InfoExtVal\(_{k} \) means larger information externalities of firm \( k \), as firm \( k \)'s ROA becomes more informative for the fundamentals of other firms in the industry.

\[
InfoExtVal_{k} = \ln \left[ \frac{PCORR_{ROA}_{k}}{1 - PCORR_{ROA}_{k}} \right] \quad (4)
\]

3.3. Information externalities and research resource allocation

We run regression (5) to test whether analysts allocate their research effort according to information externalities.

\[
Allocation_{i,t} = \alpha + \beta InfoExtVal_{i,t-1} + \delta X_{i,t-1} + IND-YEAR + \epsilon_{i,t} \quad (5)
\]

In model (5), subscript \( i \) denotes the firm and subscript \( t \) denotes the year. We measure the dependent variable Allocation for firm \( i \) in year \( t \) in different ways: the number of analysts following (Coverage\(_{i,t} \)), the number of new analysts following (NewCov\(_{i,t} \)), the number of reports issued by analysts (Rpt\(_{i,t} \)), the number of earnings revisions made by analysts (Rev\(_{i,t} \)), and the number of site visits by analysts (SiteVisit\(_{i,t} \)). As the site visit data are only
available for firms listed on the Shenzhen Stock Exchange (SZSE) after 2012, the corresponding sample where SiteVisit\textsubscript{\textit{i,t}} is the dependent variable contains only SZSE-listed firms, and the sample period is from 2012 to 2018.

In model (5), \textit{X} represents a set of control variables. Following Lang and Lundholm (1996) and Hameed et al. (2015), we control for firm size (\textit{Mv}), book-to-market ratio (\textit{Btm}), return on equity (\textit{Roe}), institutional investors’ shareholding ratio (\textit{InsHold}), stock turnover (\textit{Turnover}), and stock volatility (\textit{RetStd}). We also control for \textit{IndStatus}, the average of the relative ranking of a firm’s size, revenue, and ROA within the industry (from 1 to 5, in ascending order), to ensure that \textit{InfoExtVal} measures a firm’s information externalities beyond its position in the industry. We also include \textit{IND-YEAR}, two-dimensional higher-order fixed effects, to ensure that all parameters are estimated using variations within intersections of industries and years. The standard errors of regression (5) are clustered by industry, and we expect \textit{\beta} to be significantly positive.

We also conduct analyst-level analysis to further examine how individual analysts allocate their time and effort. In regression (6), subscripts \textit{i}, \textit{j}, and \textit{t} denote firm \textit{i}, analyst \textit{j}, and year \textit{t}, respectively. By controlling for high-dimensional fixed effects (\textit{ANALYST-YEAR}) \textit{, \beta} estimates how individual analysts allocate their time and effort within their stock portfolios.\footnote{There is no need to control for analyst characteristics because we add ANALYST-YEAR fixed effects in regression (6). Both time-varying analyst characteristics, such as experience, and time-invariant analyst characteristics, such as gender and professional background, are absorbed by the ANALYST-YEAR fixed effects.}

As in regression (5), in regression (6) we measure analyst \textit{j}’s effort allocated to stock \textit{i} in year \textit{t} by the number of reports (\textit{AnaRpt}_{\textit{i,j,t}}) and the number of earnings revisions (\textit{AnaRevision}_{\textit{i,j,t}}) made by analyst \textit{j}. In addition, we consider the number of in-depth reports issued by analyst \textit{j} that forecast both earnings and revenue (\textit{Thoroughness}_{\textit{i,j,t}}) and analyst \textit{j}’s response timeliness to firm \textit{i}’s annual report (\textit{ResponseDays}_{\textit{i,j,t}}), that is, the number of days between firm \textit{i}’s annual earnings announcement and analyst \textit{j}’s update of their earnings forecasts. Finally, we use the magnitude of the market reaction to analysts’ earnings forecast revisions (\textit{MktReaction}) as an indirect measure of analyst effort allocation. Our logic here is that more effort leads to higher-quality reports and therefore greater price impacts on the stock market. We measure \textit{MktReaction} as the cumulative excess return within the [−1, +1] analyst earnings forecast revision window.\footnote{Analysts typically cover stocks in the same industry; we find that the inclusion or exclusion of high-dimensional fixed effects \textit{IND-YEAR} does not change the main results.}

\begin{equation}
\text{Allocation}_{\textit{i,j,t}} = \alpha + \beta \text{InfoExtVal}_{\textit{i,t−1}} + \delta' X_{\textit{i,t−1}} + \text{ANALYST-YEAR} + \text{IND-YEAR} + \epsilon_{\textit{i,j,t}} \tag{6}
\end{equation}

In regressions (5) and (6), both \textit{InfoExtVal} and the control variables are lagged by 1 year relative to the dependent variable. To ensure the robustness of our empirical results, the standard errors of regression (6) are double-clustered by analyst and by firm. Furthermore, to mitigate the impact of outliers, all continuous variables used are winsorised at the 1% and 99% levels.
3.4. Information externalities and pricing efficiency

Analysts and other participants in the stock market optimise their limited research resources by allocating more time and effort to stocks with higher information externalities, which in turn could have a significant effect on pricing efficiency. We expect industry information to be rapidly incorporated into the price of stocks with high information externalities but with a delay for stocks with low information externalities. To test the above inferences, we use time series and multivariate cross-sectional regression models.

Following Hou (2007), for each year and industry, we build a High_InfoExtVal portfolio containing stocks with InfoExtVal above the 70th percentile of the industry, and a Low_InfoExtVal portfolio containing stocks with InfoExtVal below the 30th percentile of the industry. We compute daily portfolio returns for the High_InfoExtVal and Low_InfoExtVal portfolios for each industry from 1 May 2007 to 30 April 2019. Then, we test whether there is a significant lead-lag effect between the High_InfoExtVal and Low_InfoExtVal portfolios using a vector autoregressive (VAR) model, as shown in equation systems (7).

\[
\text{RetH}_{\text{IND},t} = a_0 + \sum_{k=1}^{K} a_k \text{RetH}_{\text{IND},t-k} + \sum_{k=1}^{K} b_k \text{RetL}_{\text{IND},t-k} + e_{H,\text{IND},t} \quad (7-1)
\]

\[
\text{RetL}_{\text{IND},t} = c_0 + \sum_{k=1}^{K} c_k \text{RetH}_{\text{IND},t-k} + \sum_{k=1}^{K} d_k \text{RetL}_{\text{IND},t-k} + e_{L,\text{IND},t} \quad (7-2)
\]

where \( H \) and \( L \) stand for High_InfoExtVal and Low_InfoExtVal portfolios, respectively. The subscript IND stands for industry, and the \( K \) above the summation symbol represents the highest order of time lag. To improve robustness, we report the results for both the first-order and fifth-order VAR models.

The time series model depicts the diffusion of industry information from one category of stocks to another category. According to the information delay hypothesis, industry information should first transmit to the price of stocks with high InfoExtVal and then to the price of stocks with low InfoExtVal with a time lag. That is, in the VAR model of equation (7), the lagged returns of portfolio \( H, \text{RetH}_{\text{IND},t-k} \), can significantly predict the current return of portfolio \( L, \text{RetL}_{\text{IND},t} \). However, the lagged return of portfolio \( L, \text{RetL}_{\text{IND},t-k} \), should not have significant return prediction ability for portfolio \( H \), that is, \( c_1 + \ldots + c_k \geq b_1 + \ldots b_k \).

Similar to Hou (2007), we estimate the VAR models jointly across all industries.

Following Hou and Moskowitz (2005), Kong et al. (2013), and Hu et al. (2016), we also construct IndDelay, a series of information delay measures for individual stocks.

\[
r_{i,t} = a_i + \beta_0 r_{M,t} + \sum_{k=1}^{5} \beta_k r_{M,t-k} + \delta_0 r_{\text{IND},t} + \sum_{k=1}^{5} \delta_k r_{\text{IND},t-k} + \epsilon_{i,t} \quad (8)
\]

where \( r_{i,t} \) is the daily stock return of firm \( i \), \( r_{M,t} \) is the market return, and \( r_{\text{IND},t} \) is the portfolio return of the industry to which firm \( i \) belongs. At the end of April of each year \( t + 1 \), we regress each stock’s daily returns on the market returns of same day and the last 5 days, and the corresponding industry portfolios from May of year \( t \) to April of year \( t + 1 \).
The first individual stock delay measure \( \text{IndDelay1} \) is the fraction of returns explained by lagged industry portfolio returns. As shown in equation (9), \( \text{IndDelay1} \) is 1 minus the ratio of the \( R^2 \) of regression (8) to the \( R^2 \) of a restricted regression, where \( \delta_k \) is restricted to 0 for all \( k \in [1, 5] \).

\[
\text{IndDelay1} = 1 - \frac{R^2_{\delta_k = 0, \forall k \in [1, 5]}}{R^2} \quad (9)
\]

The second information delay measure \( \text{IndDelay2} \) is given in equation (10), which is the ratio of five lagged estimates to the sum of current and lagged estimates.

\[
\text{IndDelay2} = \frac{\sum_{k=1}^{5} \delta_k}{\delta_0 + \sum_{k=1}^{5} \delta_k} \quad (10)
\]

As shown in equation (11), relative to \( \text{IndDelay2} \), the third delay measure \( \text{IndDelay3} \) further incorporates the precision of the estimates, where \( \text{se}(\delta_k) \) denotes the standard error of \( \delta_k \). \( \text{IndDelay1} \), \( \text{IndDelay2} \), and \( \text{IndDelay3} \) are higher with a slower diffusion of industry information.

\[
\text{IndDelay3} = \frac{\sum_{k=1}^{5} \frac{\delta_k}{\text{se}(\delta_k)}}{\frac{\delta_0}{\text{se}(\delta_0)} + \sum_{k=1}^{5} \frac{\delta_k}{\text{se}(\delta_k)}} \quad (11)
\]

We test whether \( \text{InfoExtVal} \) is systematically associated with \( \text{IndDelay} \) in multivariate cross-sectional regression (12). Compared with the time series test, the advantage of the regression method is that we can control for multiple variables related to information delay simultaneously. If stocks with higher \( \text{InfoExtVal} \) are priced more efficiently with less information delay, the regression coefficient \( \beta \) is expected to be significantly negative.

\[
\text{IndDelay}_{i,t} = \alpha + \beta \text{InfoExtVal}_{i,t} + \delta X_{i,t} + \text{IND-YEAR} + \epsilon_{i,t} \quad (12)
\]

Consistent with regression (5), we include high-dimensional \( \text{IND-YEAR} \) fixed effects, cluster standard errors by industry, and winsorise all continuous variables at the 1st and 99th percentiles. Table 1 reports the detailed definitions of all variables used in empirical analysis.

4. Main empirical results

4.1. Descriptive statistics

Table 2 reports the descriptive statistics for the main variables. It shows that the average incremental explanatory power of a firm’s ROA on the ROA of other firms in the same industry is about 27%. The averages for \( \text{Coverage} \), \( \text{Rpt} \), and \( \text{Revision} \) are 14, 27, and 14, respectively. The distributions of \( \text{Coverage} \), \( \text{Rpt} \), and \( \text{Revision} \) are all right-skewed, as all of their medians are lower than their means. For firms listed on SZSE that disclose information on firm–investor relationships, they receive an average of 7 analyst site visits per year,
Table 1. Variable definitions.

| Variable                        | Definition                                                                                                                                 |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| **Information Externalities**   |                                                                                                                                              |
| InfoExtVal_{i,t-1}              | Industry-level information externality measure following Hameed et al. (2015). The construction of InfoExtVal is described in detail in Section 3.2 |
| ProInfoExtVal_{i,t-1}           | Geographic region-level information externality measure constructed like InfoExtVal.                                                      |
| **Analyst Research Resource Allocation** |                                                                                                                                 |
| Coverage_{i,t}                  | The natural logarithm of (1 + the number of analysts following firm i in year t)                                                             |
| NewCov_{i,t}                    | The natural logarithm of (1 + the number of new analysts following firm i in year t)                                                         |
| Rpt_{i,t}                       | The natural logarithm of (1 + the number of reports issued by analysts following firm i in year t)                                          |
| Revision_{i,t}                  | The natural logarithm of (1 + the number of earnings forecast revisions issued by analysts following firm i in year t)                     |
| SiteVisit_{i,t}                 | The natural logarithm of (1 + the number of site visits by analysts for firm i in year t)                                                   |
| AnaRpt_{i,j,t}                  | The natural logarithm of (1 + the number of reports issued by analyst j for firm i in year t)                                              |
| AnaRevision_{i,j,t}             | The natural logarithm of (1 + the number of earnings forecast revisions issued by analyst j for firm i in year t)                          |
| ResponseDays_{i,t}              | The natural logarithm of (1 + the number of days between the announcement of firm i's annual report and analyst j's update of their earnings forecast for firm i) |
| Thoroughness_{i,t}              | The natural logarithm of (1 + the number of in-depth reports containing both revenue and earnings forecasts issued by analyst j for firm i in year t) |
| IndDelay1                       | Industry information delay for individual stocks constructed using equations (8) and (9)                                                |
| IndDelay2                       | Industry information delay for individual stocks constructed using equations (8) and (10)                                                 |
| IndDelay3                       | Industry information delay for individual stocks constructed using equations (8) and (11)                                                 |
| ProvDelay1                      | Province information delay for individual stocks, constructed like IndDelay1                                                              |
| ProvDelay2                      | Province information delay for individual stocks, constructed like IndDelay2                                                              |
| ProvDelay3                      | Province information delay for individual stock, constructed like IndDelay3                                                               |
| **Spillover Effects of Earnings Announcements** |                                                                                                                                 |
| AvgPeerRet_{i,t}                | The average of other firms' market-adjusted cumulative excess returns in the industry, during firm i's annual earnings announcement window [-1, +1] in year t |
| UE_{i,t}                        | Firm i's earnings in year t minus its earnings in year t-1, divided by total market capitalisation at the end of year t-1                  |
| AdjRet_{i,t}                    | Firm i's market-adjusted cumulative excess stock returns during the annual earnings announcement window [-1, +1] in year t                |
| **Research Resource Allocation for Professional Investors** |                                                                                                                                 |
| FundVisit_{i,t}                 | The natural logarithm of (1 + the number of site visits by mutual funds to firm i in year t)                                              |
| OtherVisit_{i,t}                | The natural logarithm of (1 + the number of site visits by other institutional investors to firm i in year t)                               |
| **Supply Chains and Exogenous Shocks** |                                                                                                                                 |
| SupplyChainPO                   | A variable with a value of 1 for firms in the experimental group and 0 for firms in the control group. A firm is assigned to the experimental group if its customers or suppliers have gone public within the 6-year window |

(Continued)
| Variable       | Definition                                                                                                                                                                                                 |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PostIPO        | A variable with a value of 1 for the experimental firm and its matched control firm after the IPO of suppliers or customers of the experimental firm and 0 before the IPO                                                                 |
| Control Variables                                                                                      |                                                                                                                                                   |
| $M_{i,t-1}$    | The natural logarithm of the total market value of outstanding shares at the end of year $t-1$                                                                                                              |
| $Btm_{i,t-1}$  | Book-to-market ratio, calculated as the book value of equity at the end of year $t-1$ divided by the total market value of outstanding shares at the end of year $t-1$                                             |
| $Roe_{i,t-1}$  | Return on equity, calculated as the net income in year $t-1$ divided by the book value of equity at the end of year $t-1$                                                                                     |
| $InsHold_{i,t-1}$ | The shareholding ratio of institutional investors at the end of year $t-1$                                                                                                                                |
| $Turnover_{i,t-1}$ | The average daily turnover ratio of shares in year $t-1$                                                                                                                                             |
| $RetStd_{i,t-1}$ | The standard deviation of daily stock returns in year $t-1$                                                                                                                                          |
| $IndStatus_{i,t-1}$ | Firm $i$'s industry position in year $t-1$. For each year, the firms in each industry are sorted into quintiles in ascending order according to their size, revenue, and ROA. $IndStatus_{i,t-1}$ is the average of the relative rankings of size, revenue, and ROA. |
| $ProStatus_{i,t-1}$ | Firm $i$'s geographic economic position in year $t-1$. For each year, the firms in each province are sorted into quintiles in ascending order according to their size, revenue, and ROA. $ProStatus_{i,t-1}$ is the average of the relative rankings of size, revenue, and ROA. |
but the median of AnaVisit is 1, indicating that a few firms receive more attention from analysts. The descriptive statistics for the other variables are generally in line with expectations.

Table 3 compares the characteristics of firms with different InfoExtVal. For each year, within each industry, we sort stocks into quintiles based on InfoExtVal. Quintile 5 represents stocks with the highest InfoExtVal, whereas Quintile 1 represents stocks with the lowest InfoExtVal. We calculate the mean and median values of Mv (in millions of RMB), Btm, and Roe for all stocks in each quintile, as well as the ranking of the above variables (from 5 to 1, in descending order).

| Variable       | Mean      | P25       | Median   | P75       | Std Deviation |
|----------------|-----------|-----------|----------|-----------|---------------|
| InfoExtVal     | 0.2709    | 0.1457    | 0.2618   | 0.3848    | 0.1384        |
| Coverage       | 13.5784   | 1.0000    | 7.0000   | 21.0000   | 16.3668       |
| NewCov         | 10.0619   | 1.0000    | 5.0000   | 16.0000   | 11.8741       |
| Rpt            | 27.2031   | 1.0000    | 11.0000  | 38.0000   | 37.5574       |
| Revision       | 13.5936   | 0.0000    | 3.0000   | 18.0000   | 21.9990       |
| AnaVisit       | 8.7913    | 0.0000    | 3.0000   | 12.0000   | 14.0320       |
| AnaRpt         | 2.0203    | 1.0000    | 1.0000   | 2.0000    | 1.7963        |
| ResponseDays   | 130.4491  | 20.0000   | 118.0000 | 217.0000  | 111.5497      |
| Thoroughness   | 0.9023    | 0.0000    | 0.0000   | 1.0000    | 1.3973        |
| MktReaction    | 0.0067    | −0.0206   | 0.0008   | 0.0273    | 0.0689        |
| Freq           | −0.0641   | −0.0526   | 0.0000   | 0.0000    | 2.3149        |
| IndDelay1      | 0.0431    | 0.0136    | 0.0271   | 0.0529    | 0.0477        |
| IndDelay2      | 0.0990    | −0.1838   | 0.1431   | 0.4085    | 1.7281        |
| IndDelay3      | 0.0929    | −0.1829   | 0.1412   | 0.4058    | 1.6967        |
| Mv             | 22.5015   | 21.8172   | 22.4341  | 23.0937   | 0.9803        |
| Btm            | 0.3574    | 0.1849    | 0.3037   | 0.4738    | 0.2427        |
| Roe            | 0.0570    | 0.0232    | 0.0644   | 0.1149    | 0.1731        |
| InstHold       | 0.3888    | 0.1893    | 0.3962   | 0.5728    | 0.2358        |
| Turnover       | 0.0189    | 0.0093    | 0.0157   | 0.0254    | 0.0126        |
| RetStd         | 0.0311    | 0.0235    | 0.0290   | 0.0375    | 0.0103        |
| IndStatus      | 3.0396    | 2.3333    | 3.0000   | 4.0000    | 1.0676        |
| ProStatus      | 3.0498    | 2.3333    | 3.0000   | 4.0000    | 1.0486        |

The descriptive statistics for InfoExtVal to AnaVisit in Table 2 are not log transformed.
Table 4. Information externalities and research resource allocation: firm-level analysis.

| Variable | (1) Coverage | (2) NewCov | (3) Rpt | (4) Revision | (5) SiteVisit |
|----------|--------------|------------|--------|-------------|--------------|
| InfoExtVal | 0.3179*** | 0.2823*** | 0.3740*** | 0.3109*** | 0.2780*** |
|          | (10.5774)  | (10.5499)  | (9.7882) | (8.3829)    | (7.3170)     |
| Mv       | 0.4467*** | 0.3949*** | 0.5431*** | 0.5402***   | 0.3121***    |
|          | (11.5804)  | (11.1456)  | (11.7254) | (12.7018)   | (5.6667)     |
| Btm      | -0.3031*** | -0.3007*** | -0.3680*** | -0.5049***  | -0.3199***   |
|          | (-3.0340)  | (-3.3254)  | (-3.0874) | (-4.0577)   | (-2.8519)    |
| Roe      | 0.5353*** | 0.5018*** | 0.6574*** | 0.5995***   | 0.7156***    |
|          | (6.1133)   | (6.4932)   | (6.0552) | (5.2557)    | (4.0757)     |
| InsHold  | 0.4989*** | 0.4386*** | 0.6258*** | 0.6444***   | 0.0083       |
|          | (7.9177)   | (8.5731)   | (8.0564) | (8.5195)    | (0.0564)     |
| Turnover | -0.3636    | -0.4631   | -1.4294  | -3.1415*    | -4.2644      |
|          | (-0.2051)  | (-0.2728)  | (-0.7072) | (-1.9264)   | (-1.4918)    |
| RetStd   | 1.9407     | 2.4587    | 1.9662   | -2.7384     | 3.6288       |
|          | (0.4731)   | (0.6297)   | (0.4186) | (-0.6698)   | (1.2807)     |
| IndStatus| 0.3864***  | 0.3459*** | 0.4676*** | 0.3968***   | 0.1956***    |
|          | (11.7341)  | (11.7778)  | (11.5736) | (10.6774)   | (7.8250)     |
| IND-YEAR | Yes        | Yes       | Yes      | Yes         | Yes          |
| Adj-R²   | 0.5241     | 0.4964    | 0.5227   | 0.4939      | 0.2477       |
| N        | 19,739     | 19,739    | 19,739   | 19,739      | 8,630        |

This table reports the regression coefficients and corresponding t-statistics in parentheses. Standard errors are clustered by analyst. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5. Information externalities and research resource allocation: analyst-level analysis.

| Variable | (1) AnaRpt | (2) AnaRevision | (3) Thoroughness | (4) ResponseDays | (5) MktReaction |
|----------|------------|-----------------|------------------|------------------|----------------|
| InfoExtVal | 0.0098*** | 0.0146*** | 0.0050* | -0.0462*** | -0.0153 |
|          | (3.7575)  | (3.8835) | (1.8667) | (-6.7068)    | (-0.4288)     |
| InfoExtVal × Frev | 0.0235** | (1.9954) | 0.0539*** | | |
| Frev      | | | | | |
|          | | | | | |
| Mv       | 0.0443*** | 0.0631*** | 0.0179*** | -0.0934*** | -0.1137*** |
|          | (11.3937) | (11.4822) | (4.1864) | (-11.5542)  | (-3.0389)    |
| Btm      | -0.0301*** | -0.0422*** | -0.0396*** | 0.0553***   | 0.0663       |
|          | (-2.9291) | (-2.8396) | (-3.4474) | (2.0547)     | (0.5226)     |
| Roe      | 0.2096*** | 0.3081*** | 0.1548*** | -0.2413***  | -0.8266**    |
|          | (9.5810)  | (9.7553) | (6.1775) | (-3.8810)   | (-2.2165)    |
| InsHold  | 0.0377*** | 0.0559*** | 0.0297*** | -0.0615***  | -0.2133*     |
|          | (4.5063)  | (4.6270) | (3.2359) | (-2.8501)   | (-1.7511)    |
| Turnover | -0.1848   | -0.2893   | -0.2483  | -0.3365     | -3.2424      |
|          | (-0.9301) | (-1.0021) | (-1.1685) | (-0.6350)   | (-0.9516)    |
| RetStd   | 0.5661*   | 0.8828*   | -0.3051  | -1.4724     | 0.7761       |
|          | (1.7247)  | (1.8478) | (-0.8645) | (-1.5470)   | (0.1419)     |
| IndStatus| 0.0106*** | 0.0170*** | 0.0074** | -0.0490***  | -0.0668      |
|          | (3.3681)  | (3.7547) | (2.1626) | (-6.3517)   | (-1.5272)    |
| ANALYST-YEAR | Yes | Yes | Yes | Yes | Yes |
| IND-YEAR  | Yes | Yes | Yes | Yes | Yes |
| Adj-R²   | 0.3484   | 0.3450   | 0.4869  | 0.5426      | 0.0436       |
| N        | 258,849  | 258,849  | 258,849 | 255,880     | 162,579      |

This table reports the regression coefficients and corresponding t-statistics in parentheses. Standard errors are clustered by analyst and firm. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.
As shown in Table 3, firms with high InfoExtVal are generally larger and have higher book-to-market ratios and stronger profitability within the industry than firms with low InfoExtVal. Table 3 highlights the need to control firm characteristics, such as Mv, Btm, and Roe, in the regression analysis.

4.2. Firm-Level tests

Table 4 reports the results of the firm-level analysis. In columns (1) and (2), we find that both Coverage and NewCov are significantly and positively associated with InfoExtVal, implying that analysts are more attentive to stocks with high information externalities than to those with low information externalities. Columns (3) and (4) further illustrate that analysts release more reports and forecast revisions for firms with high InfoExtVal. Moreover, in column (5), we find that analysts are more likely to conduct site visits to firms with high InfoExtVal than to those with low InfoExtVal.

Collectively, the empirical results in Table 4 suggest that analysts allocate more research resources to stocks with high InfoExtVal, supporting Hypothesis 1.
Table 7. Information externalities and pricing efficiency: multivariate regression tests.

| Variable     | (1)     | (2)     | (3)     |
|--------------|---------|---------|---------|
| InfoExtVal   | −0.0032*** | −0.0513*** | −0.0555*** |
|              | (−4.0415)   | (−2.4185)   | (−2.5391)   |
| Mv           | −0.0011    | −0.0421*   | −0.0362    |
|              | (−1.1253)   | (−1.9893)   | (−1.4651)   |
| Btm          | −0.0332*** | −0.2649**  | −0.2525**  |
|              | (−10.0454)  | (−2.4013)   | (−2.3368)   |
| Roe          | −0.0107*** | 0.0272     | 0.0665     |
|              | (−3.1961)   | (0.3610)    | (0.7437)    |
| InsHold      | −0.0060**  | −0.0939    | −0.0739    |
|              | (−2.6824)   | (−1.3877)   | (−1.0762)   |
| Coverage     | −0.0032*** | 0.0059     | 0.0080     |
|              | (−5.6498)   | (0.4191)    | (0.4587)    |
| Turnover     | −0.4437*** | −2.4409    | −3.1769    |
|              | (−12.9210)  | (−1.1419)   | (−1.3861)   |
| RetStd       | −0.1821    | −1.9138    | 0.9010     |
|              | (−1.5296)   | (−0.6777)   | (0.3064)    |
| IndStatus    | −0.0040*** | −0.0552*** | −0.0613*** |
|              | (−5.4409)   | (−3.6428)   | (−4.0457)   |
| IND-YEAR     | Yes       | Yes       | Yes       |
| Adj-R²       | 0.2120     | 0.0136     | 0.0134     |
| N            | 19,410     | 19,410     | 19,410     |

This table reports the regression coefficients and corresponding t-statistics in parentheses. Standard errors are clustered by industry. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

4.3. Analyst-Level tests

Although Table 4 derives firm-level evidence consistent with analysts allocating research resources based on information externalities, we cannot determine how an individual analyst allocates resources for stocks in their portfolios. In Table 5, we use an analyst-firm-year sample including ANALYST-YEAR fixed effects to examine individual analyst behaviour.

Column (1) of Table 5 shows that analysts release more reports on stocks with high (versus low) information externalities; column (2) shows that analysts also pay more attention to fundamental changes in these stocks and revise their earnings forecasts for these stocks more frequently. Furthermore, analysts are more likely to issue in-depth reports that provide both earnings and revenue forecasts for stocks with high information externalities; as shown in column (3), InfoExtVal has a significantly positive correlation with Thoroughness. In column (4), we also demonstrate that analysts respond more quickly to the earnings announcements of stocks with high information externalities, as the time lag of forecast revisions after the publication of annual reports for stocks with high information externalities is significantly shorter.

In addition, we investigate the market reactions to analyst forecast revisions. If analysts expend more resources on stocks with high information externalities, the corresponding forecast revisions should be more informative and have a greater price impact. As shown in column (5), the interaction term InfoExtVal × Frev is significantly positive, indicating that, given the magnitude of forecast revision, investors react strongly to stocks with high information externalities.
The findings of Table 5 clarify the micro-mechanisms underlying research resource allocation and further support our Hypothesis 2.

### 4.4. Information externalities and pricing efficiency

In this section, we use both time series VAR tests and cross-sectional multivariate regressions to examine the pricing efficiency of stocks with different information externalities.

In regression (1) of Table 6, RetH and RetL are lagged by one period. The results show that lagged RetH (L1.RetH) can predict RetL, but lagged RetL (L1.RetL) has no explanatory power for RetH. The corresponding F-test shows that the above asymmetry in cross-autocorrelations is significant at the 5% level.

In regression (2) of Table 6, we use a fifth-order VAR test to conduct a robustness check. We find that four of the five lagged values of RetH are significantly positive when the current RetL is the dependent variable, but none of the five lagged values of RetL can predict current RetH. The F-test rejects $\sum L. RetH = \sum L. RetL$ at the 1% significance level.
The above VAR test results show a significant lead-lag effect, indicating that
industry information is incorporated into the price of stocks with high information
externalities first and then spreads to the price of stocks with low information
externalities.

Table 7 reports the results of multivariate cross-sectional regression tests. We control
for other variables related to pricing efficiency and information delay, such as size (Mv; Hou, 2007; Lo & MacKinlay, 1990), institutional investor ownership (InstHold; Sias & Starks,
1997), analyst coverage (Coverage; Brennan et al., 1993), and stock trading volume
(Turnover; Chordia & Swaminathan, 2000).

The results of the regression tests highly collaborate with the findings in Table 6. As we
see in columns (1)–(3) of Table 7, all three individual industrial information delay measures
(IndDelay) are significantly and negatively related to InfoExtVal. In other words, stocks with
high InfoExtVal are priced more efficiently as they absorb industry information faster than
stocks with low InfoExtVal.

The results of Tables 6 and 7 also deepen our understanding of pricing efficiency. In
the classical framework of Grossman and Stiglitz (1980) and Verrecchia (1982), pricing
efficiency is determined by investors’ information acquisition incentives, which in turn
are influenced by information costs. We distinguish the information externality channel
that affects investors’ information acquisition incentives from the traditional information
cost channel. We posit that investors have greater information acquisition incentives for stocks with high information externalities. As investors expend more resources
on such stocks, they are likely to obtain information that is valuable for many other stocks.

| Variable       | (1) Coverage | (2) NewCov | (3) Rpt | (4) Revision | (5) SiteVisit |
|----------------|--------------|------------|--------|-------------|--------------|
| ProInfoExtVal  | 0.3104***    | 0.2774***  | 0.3658*** | 0.3044***   | 0.2782***    |
| (14.4901)      | (13.8108)    | (13.9704)  | (11.8931) | (9.1167)    |
| Mv             | 0.5338***    | 0.4763***  | 0.6398*** | 0.6053***   | 0.4023***    |
| (22.8834)      | (21.5464)    | (22.6778)  | (22.5645) | (7.4882)    |
| Btm            | −0.3592***   | −0.3805*** | −0.4259*** | −0.5876***  | −0.2887**    |
| (−4.6361)      | (−5.2533)    | (−4.5712)  | (−6.4028) | (−2.3991)   |
| Roe            | 0.4247***    | 0.3963***  | 0.5246*** | 0.4702***   | 0.4298***    |
| (8.0230)       | (7.7042)     | (8.1705)   | (7.2429)  | (3.6608)    |
| InsHold        | 0.5820***    | 0.4963***  | 0.7360*** | 0.7699***   | −0.1025      |
| (4.7908)       | (4.6468)     | (4.9541)   | (6.0380)  | (−1.0758)   |
| Turnover       | 0.1472       | −0.1175    | −0.9861  | −2.7486     | −1.8250      |
| (0.0728)       | (−0.0644)    | (−0.4065)  | (−1.3594) | (−0.6619)   |
| RetStd         | 1.2504       | 2.1591     | 1.3639   | −3.8317*    | −1.4466      |
| (0.7143)       | (1.2280)     | (0.6486)   | (−1.7151) | (−0.3577)   |
| ProStatus      | 0.3227***    | 0.2890***  | 0.3914*** | 0.3317***   | 0.1163***    |
| (13.4668)      | (13.0077)    | (13.6905)  | (13.5636) | (3.4925)    |
| PRO-YEAR       | Yes          | Yes        | Yes      | Yes         | Yes          |
| Adj-R²         | 0.4991       | 0.4696     | 0.4978   | 0.4686      | 0.2326       |
| N              | 14,029       | 14,029     | 14,029   | 14,029      | 5,919        |

This table reports the regression coefficients and corresponding t-statistics in parentheses. Standard errors are clustered by province. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.
Table 10. Geographic information externalities and pricing efficiency: lead-lag effect tests.

| Variable | RetL | RetH | RetL | RetH |
|----------|------|------|------|------|
|          | (1)  |      | (2)  |      |
| L1. RetL | 0.0719*** | −0.0085 | 0.0699*** | −0.0051 |
|          | (11.9137) | (−1.4786) | (11.4596) | (−0.8838) |
| L2. RetL | −0.0151** | −0.0114* | (−2.4641) | (−1.9560) |
| L3. RetL | 0.0129** | 0.0142** | (2.1077) | (2.4416) |
| L4. RetL | 0.0131** | −0.0001 | (2.1513) | (−0.0217) |
| L5. RetL | −0.0370*** | −0.0241*** | (−6.1055) | (−4.1694) |

L1. RetH 0.0706*** 0.1016***
(11.0912) (16.7663)
L2. RetH 0.0032 0.0246***
(0.4992) (−4.0324)
L3. RetH 0.0289*** 0.0174***
(4.5058) (2.8543)
L4. RetH 0.0314*** 0.0350***
(4.8987) (5.7350)
L5. RetH 0.0276*** 0.0017
(4.3125) (0.2791)

F Test 36.20*** 40.89***

This table reports the regression coefficients and t-statistics for the VAR model. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 11. Geographic information externalities and pricing efficiency: multivariate regression tests.

| Variable          | (1) ProvDelay1 | (2) ProvDelay2 | (3) ProvDelay3 |
|-------------------|----------------|----------------|----------------|
| ProInfoExtVal     | −0.0037***     | −0.0448***     | −0.0423***     |
|                   | (−6.0079)      | (−2.9873)      | (−2.7492)      |
| Mv                | −0.0000        | 0.0341         | 0.0353         |
|                   | (−0.0219)      | (0.9229)       | (1.0079)       |
| Btm               | −0.0320***     | 0.1885**       | 0.0689         |
|                   | (−13.5385)     | (2.3889)       | (0.8548)       |
| Roe               | −0.0077***     | 0.0645         | 0.0481         |
|                   | (−3.4087)      | (0.6922)       | (0.5090)       |
| InsHold           | −0.0036*       | −0.0620        | −0.1284*       |
|                   | (−1.7329)      | (−0.8157)      | (−1.7586)      |
| Coverage          | −0.0027***     | −0.0118        | 0.0010         |
|                   | (−5.7306)      | (−0.6049)      | (0.0555)       |
| Turnover          | −0.4484***     | −3.8764        | −3.9598        |
|                   | (−9.3422)      | (−1.4711)      | (−1.4994)      |
| RetStd            | −0.1295*       | −1.4610        | −1.0676        |
|                   | (−1.7571)      | (−0.3212)      | (−0.2505)      |
| ProStatus         | −0.0040***     | −0.0946***     | −0.0977***     |
|                   | (−6.4465)      | (−3.6284)      | (−4.7474)      |
| PRO-YEAR          | Yes            | Yes            | Yes            |
| Adj-R²            | 0.2011         | 0.0041         | 0.0026         |
| N                 | 13,788         | 13,788         | 13,788         |

This table reports the regression coefficients and corresponding t-statistics in parentheses. Standard errors are clustered by province. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.
5. Further analysis

5.1. Evidence from the spillover effects of earnings announcements

In this section, we test whether the market pays more attention to stocks with high information externalities by focusing on the earnings announcements of such firms. Foster (1981) and Wang (2014) point out that earnings announcements are important sources of industry information for investors, and the earnings announcements of bellwether firms usually have spillover effects on industry peers.

In Table 8, the dependent variable AvgPeerRet<sub>i,t</sub> measures the average cumulative abnormal returns relative to the market returns of firm <i>i</i>’s industry peers when firm <i>i</i> releases its annual report, and the independent variable UE<sub>i,t</sub> is earnings growth deflated by the market value of equity. Therefore, the coefficient of UE<sub>i,t</sub> estimates the spillover effect of firm <i>i</i>’s earnings announcement on other firms within the same industry. We interact High InfoExtVal with UE and expect the coefficient of High InfoExtVal × UE to be positive, as the earnings information of firms with high information externalities is also useful for other stocks in the same industry. To rule out the possibility that stock returns mechanically comove with other stocks in the same industry, we also control for IndBeta, which is the regression coefficient of daily stock returns on daily industry portfolio returns (excluding firm <i>i</i>).

Consistent with the findings of Foster (1981) and Wang (2014), in column (1), the coefficient of UE is significantly positive, suggesting that earnings announcements contain industry information and have spillover effects on the stock prices of industry peers.

### Table 12. Information externalities in supply chains.

| Variable           | (1) Coverage | (2) NewCov | (3) Rpt | (4) Revision | (5) SiteVisit |
|--------------------|--------------|------------|---------|--------------|---------------|
| SupplyChainIPO × PostIPO | 0.1074**     | 0.0846**   | 0.1167** | 0.0819       | 0.1518        |
|                    | (2.3907)     | (1.9744)   | (2.1439) | (1.4759)     | (1.2336)      |
| SupplyChainIPO     | 0.0248       | 0.0164     | 0.0546  | 0.0812       | −0.1882**     |
|                    | (0.5170)     | (0.3860)   | (0.9165) | (1.4064)     | (−2.1592)     |
| PostIPO            | −0.1274***   | −0.1111*** | −0.1368*** | −0.0903*      | −0.2141*      |
|                    | (−3.4292)    | (−3.2433)  | (−2.9518) | (−1.9465)    | (−1.9360)     |
| Mv                 | 0.7282***    | 0.6793***  | 0.8856*** | 0.7386***     | 0.4310***     |
|                    | (16.0864)    | (15.8768)  | (15.9326) | (13.4887)    | (4.1529)      |
| Btm                | 0.4388***    | 0.3261***  | 0.3896*** | −0.0965      | 0.1682        |
|                    | (4.1811)     | (3.2124)   | (3.0076)  | (−0.7273)    | (0.5923)      |
| Roe                | 0.2651       | 0.3569*    | 0.3852  | 0.5606**     | 0.0020        |
|                    | (1.2935)     | (1.8291)   | (1.5147)  | (2.1427)     | (1.2797)      |
| InsHold            | 0.4357***    | 0.4311***  | 0.5567*** | 0.4495***     | 0.5951*       |
|                    | (4.2914)     | (4.3162)   | (4.4454)  | (3.3128)     | (1.8424)      |
| Turnover           | 3.7217**     | 3.1484*    | 2.3487  | −2.615       | −2.3422       |
|                    | (2.0098)     | (1.7175)   | (1.0464)  | (−1.2412)    | (−0.4648)     |
| RetStd             | 2.4427       | 5.3250*    | 3.1562  | 3.5576       | 7.0537        |
|                    | (0.7572)     | (1.7006)   | (0.8248)  | (0.9962)     | (0.8411)      |
| IndStatus          | 0.2234***    | 0.1791***  | 0.2657*** | 0.2437***     | 0.1150        |
|                    | (7.8087)     | (6.5618)   | (7.5878)  | (7.0187)     | (1.3572)      |
| FIRM               | Yes          | Yes        | Yes     | Yes          | Yes           |
| YEAR               | Yes          | Yes        | Yes     | Yes          | Yes           |
| Adj-R<sup>2</sup>  | 0.8205       | 0.7830     | 0.8222  | 0.7850       | 0.5547        |
| N                  | 4,353        | 4,353      | 4,353   | 4,353        | 1,020         |

This table reports the regression coefficients and corresponding t-statistics in parentheses. Standard errors are clustered by firm. ***, *, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.
The coefficient of \( \text{High InfoExtVal} \times UE \) is also significantly positive, indicating that the spillover effects of stocks whose fundamentals are closely related to those of their industry peers are much stronger than that of stocks whose fundamentals are weakly correlated. This finding validates the measurement of \( \text{InfoExtVal} \) and further suggests that the market values the information released by high \( \text{InfoExtVal} \) stocks and transfers that information to other stocks in the same industry.

In column (2), we use the abnormal returns of firm \( i \) during the \([-1, +1]\) earnings announcement window \((\text{AdjRet}_{i,t})\) as an alternative measure of firm \( i \)'s earnings information and find similar results.

### 5.2. Information externalities within geographic region

So far, we have empirically verified information externalities within industries; however, this does not mean that our theory is only restricted to industries. If the information externality hypothesis was generalisable, we should obtain similar findings for stocks connected by other economic links. Coval and Moskowitz (1999) show that investors have strong geographic considerations and exhibit home bias in portfolio decisions. Parsons et al. (2020) suggest that there are close economic ties between firms within the same geographic region. Therefore, to test the external validity of our theory, we consider information externalities within geographic regions.

Following the construction of \( \text{InfoExtVal} \), we define a geographic information externality measure, \( \text{ProInfoExtVal} \), within provinces. Table 9 reports the results of firm-level tests. In Table 9, the coefficients of \( \text{ProInfoExtVal} \) are significantly positive in all columns, which is highly consistent with the results in Table 3. The findings of Table 9 indicate that firms with higher geographic information externalities receive significantly more resources from analysts, which supports our theory from a different perspective.

Tables 10 and 11 examine the pricing implications of information externalities in terms of geographic regions. In Table 10, we find lead-lag effects between stocks with high and low province information externalities similar to Table 6. \( \text{RetH} \) (stocks with high \( \text{ProInfoExtVal} \)) significantly leads \( \text{RetL} \) (stocks with low \( \text{ProInfoExtVal} \)); in other words, stocks with high \( \text{ProInfoExtVal} \) respond more rapidly to province information than stocks with low \( \text{ProInfoExtVal} \).

The results of cross-sectional regression tests are reported in Table 11. We document significantly negative associations between \( \text{ProInfoExtVal} \) and \( \text{ProvDelay} \) after controlling for other factors that are likely to affect pricing efficiency, and the results are stable for all three measures of \( \text{ProvDelay} \).

---

4Due to the very high number of listed firms in Beijing, Shanghai, and Shenzhen, we exclude listed firms registered in Beijing, Shanghai, and Guangdong province in all geographic tests. Untabulated results show that our main results are not affected by this sample selection restriction.

5The analyst profession is organised by industry sector, and it is very rare for analysts to specialise by location (Parsons et al., 2020). As the portfolio choice of an analyst is not based on geographic regions, it is not meaningful to conduct individual analyst-level tests.
Table 13. Information externalities and other market participants.

|                | FundVisit | OtherVisit |
|----------------|-----------|------------|
| **InfoExtVal** | **0.2140*** | **0.2565*** |
|                | (6.0231)  | (7.0074)   |
| **Mv**         | **0.3288*** | 0.4446***  |
|                | (5.6367)  | (8.7893)   |
| **Btm**        | −0.3780*** | −0.3305*** |
|                | (−4.2834) | (−3.9372)  |
| **Roe**        | 0.7174***  | 0.6911***  |
|                | (3.8637)  | (3.2857)   |
| **Inshold**    | 0.1604    | 0.0476     |
|                | (1.1954)  | (0.3091)   |
| **Coverage**   | −4.0899   | −4.0861    |
|                | (−1.5838) | (−1.4397)  |
| **Turnover**   | 3.2218    | 2.4334     |
|                | (1.0342)  | (0.6565)   |
| **RetStd**     | 0.1690***  | 0.1758***  |
|                | (7.2649)  | (6.7638)   |
| **IndStatus**  | 0.2140***  | 0.2565***  |
|                | (6.0231)  | (7.0074)   |
| **IND-YEAR**   | Yes       | Yes        |
| **Adj-R²**     | 0.2404    | 0.2559     |
| **N**          | 8,630     | 8,630      |

This table reports the regression coefficients and corresponding t-statistics in parentheses. Standard errors are clustered by industry. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

5.3. Information externalities in supply chains

The supply chain is another type of economic link among firms (Cohen & Frazzini, 2008). If a firm’s (for convenience, let us call it the focal firm) main suppliers or customers go public, its information externalities will significantly increase after the IPO, as investors can benefit from the focal firm’s information externalities by trading the stocks of its suppliers or customers. In general, the focal firm does not determine the IPO decisions of its suppliers or customers, thus the change in information externalities caused by the IPO of its suppliers or customers will be exogenous. Therefore, we consider IPOs in the supply chain as exogenous shocks to rule out endogeneity and provide further support for our hypotheses.

We collect data on the top five customers and suppliers of A-share listed firms and conduct difference-in-differences tests. The corresponding results are reported in Table 12. SupplyChainIPO is a dummy variable with a value of 1 for the treatment group and 0 for the control group. A listed firm is assigned to the treatment group if its suppliers or customers went public during the 6-year testing window. As IPOs in supply chains are relatively rare, for each firm in the treatment group and for the year before the IPO, we match it with a control firm with comparable characteristics, including size, book-to-market ratio, return on equity, institutional shareholding, turnover ratio, and industry dummies, using propensity score matching. We compare the differences between the treatment and control groups 3 years before and 3 years after the exogenous IPO events. PostIPO takes a value of 1 for the post-IPO period and 0 for the pre-IPO period. The empirical results show that the interaction term SupplyChainIPO × Post is significantly
positive in columns (1), (2), and (3). In columns (4) and (5), the sign of SupplyChainIPO × Post is positive, and the corresponding t-statistics are close to the conventional 10% significance level.

The findings in Table 12 demonstrate the significant impact of information externalities on analysts’ research resource allocation through exogenous shocks and extend the validity of the information externality hypothesis to supply chains.

5.4. Information externalities and other stock market participants

As mentioned before, an essential prerequisite for making profits from information externalities is that investors have sufficient funds to trade a large number of stocks. Therefore, our information externality hypothesis should also apply for institutional investors, as they have strong research capabilities and abundant capital. However, unlike analysts, due to limited data, the allocation of resources by institutional investors is largely unobservable to researchers. We use site visit information disclosed by SZSE-listed firms to examine whether our hypotheses hold for mutual funds and other institutional investors. Table 13 reports the corresponding results. We find that institutional investors are more likely to conduct site visits to firms with high InfoExtVal than to those with low InfoExtVal. Table 13 provides some insights into whether institutional investors allocate their research effort based on information externalities, as site visits are inadequate to paint a complete picture of institutional investor behaviour.

6. Conclusion

In stock markets, it is important to optimise the allocation of research resources to collect value-relevant information, as such resources are limited for both investors and financial intermediaries, such as analysts. Our study focuses on the importance of information externalities, as information that is useful for many stocks is more valuable than information specific to a single stock.

We use the correlation between a firm’s financial fundamentals (ROA) and those of its industry peers as the measure of information externalities, and find that analysts are aware of the importance of information externalities and allocate research resources accordingly. Overall, we find that firms with high information externalities are associated with more analyst coverage, more reports, more forecast revisions, and more site visits than firms with low information externalities. From analyst-level tests, we also find that within stock portfolios, analysts devote more effort to stocks with high information externalities than to those with low information externalities, which is reflected in more (in-depth) analyst reports and forecast revisions for such stocks and a faster response to the earnings announcements of such stocks. We demonstrate that analyst forecast revisions are more informative and credible for stocks with high information externalities, which is also consistent with analysts’ information externality-oriented research resource allocation strategy.

We further examine the pricing implications of information externalities, showing that stocks with different information externalities have different information transmission efficiency. Stocks with high information externalities transmit industry (geographic) information rapidly, while stocks with low information externalities adjust to that information
with marked delays. The above results extend the theoretical framework of Grossman and Stiglitz (1980) and Verrecchia (1982) by demonstrating that information externalities constitute an influential factor in investors’ information acquisition incentives.

**Acknowledgement**

We are grateful for the detailed and constructive suggestions from the anonymous reviewers. We acknowledge financial supports from the Fundamental Research Funds for the Central Universities in University of International Business and Economics (19QD05) and in Zhongnan University of Economics and Law (2722022BQ035), the Tsinghua University School of Economics and Management Research Grant (2020051009), and the National Natural Science Foundation of China (72002032, 72102231 and 71672098)

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**References**

Barber, B.M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies, 21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Ben-Rephael, A., Da, Z., & Israelsen, R.D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies, 30*(9), 3009–3047. https://doi.org/10.1093/rfs/hhx031

Brennan, M.J., Jegadeesh, N., & Swaminathan, B. (1993). Investment analysis and the adjustment of stock prices to common information. *Review of Financial Studies, 6*(4), 799–824. https://doi.org/10.1093/rfs/6.4.799

Chordia, T., & Swaminathan, B. (2000). Trading volume and cross-autocorrelations in stock returns. *Journal of Finance, 55*(2), 913–935. https://doi.org/10.1111/0022-1082.00231

Clement, M.B. (1999). Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics, 27*(3), 285–303. https://doi.org/10.1016/S0165-4101(99)00013-0

Cohen, L., & Frazzini, A. (2008). Economic links and predictable returns. *Journal of Finance, 63*(4), 1977–2011. https://doi.org/10.1111/j.1540-6261.2008.01379.x

Coval, J.D., & Moskowitz, T.J. (1999). Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance, 54*(6), 2045–2073. https://doi.org/10.1111/0022-1082.00181

DellaVigna, S., & Pollet, J.M. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance, 64*(2), 709–749. https://doi.org/10.1111/j.1540-6261.2009.01447.x

Derrien, F., & Kecskés, A. (2013). The real effects of financial shocks: Evidence from exogenous changes in analyst coverage. *Journal of Finance, 68*(4), 1407–1440. https://doi.org/10.1111/jofi.12042

Driskill, M., Kirk, M.P., & Tucker, J.W. (2020). Concurrent earnings announcements and analysts. *Information Production. Accounting Review, 95*(1), 165–189. https://doi.org/10.2308/accr-52489

Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Foster, G. (1981). Intra-industry information transfers associated with earnings releases. *Journal of Accounting and Economics, 3*(3), 201–232. https://doi.org/10.1016/0165-4101(81)90003-3

Grossman, S.J., & Stiglitz, J.E. (1980). On the impossibility of informationally efficient markets. *American Economic Review, 70*(3), 393–408. https://doi.org/10.7916/D876SR99
Hameed, A., Morck, R., Shen, J., & Yeung, B. (2015). Information, analysts, and stock return comovement. *Review of Financial Studies*, 28(11), 3153–3187. https://doi.org/10.1093/rfs/hhv042

Harford, J., Jiang, F., Wang, R., & Xie, F. (2019). Analyst career concerns, effort allocation, and firms’ information environment. *Review of Financial Studies*, 32(6), 2179–2224. https://doi.org/10.1093/rfs/hhy101

Hirshleifer, D., Lim, S.S., & Teoh, S.H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance*, 64(5), 2289–2325. https://doi.org/10.1111/j.1540-6261.2009.01501.x

Hou, K., & Moskowitz, T.J. (2005). Market frictions, price delay, and the cross-section of expected returns. *Review of Financial Studies*, 18(3), 981–1020. https://doi.org/10.1093/rfs/hhi023

Hou, K. (2007). Industry information diffusion and the lead-lag effect in stock returns. *Review of Financial Studies*, 20(4), 1113–1138. https://doi.org/10.1093/revfin/hhm003

Hu, C., Li, J., & Liu, P. (2016). Price delay, speculative demand and stock returns. *Management World*, (1), 44–53. (In Chinese). https://doi.org/10.19744/j.cnki.11-1235/f.2016.01.005

Huang, S., Huang, Y., & Lin, T.-C. (2019). Attention allocation and return comovement: Evidence from repeated natural experiments. *Journal of Financial Economics*, 132(2), 369–383. https://doi.org/10.1016/j.jfineco.2018.10.006

Kacperczyk, M., Van Nieuwerburgh, S., & Veldkamp, L. (2014). Time-Varying fund manager skill. *Journal of Finance*, 69(4), 1455–1484. https://doi.org/10.1111/jofi.12084

Kacperczyk, M., Van Nieuwerburgh, S., & Veldkamp, L. (2016). A rational theory of mutual funds. *Attention Allocation*. *Econometrica*, 84(2), 571–626. https://doi.org/10.3982/ECTA11412

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kempf, E., Manconi, A., & Spalt, O. (2017). Distracted shareholders and corporate actions. *Review of Financial Studies*, 30(5), 1660–1695. https://doi.org/10.1093/rfs/hhw082

Kong, D., Liu, S., & Ying, Q. (2013). The role of media in corporate governance. *Management World*, (7), 145–162. (In Chinese). https://doi.org/10.19744/j.cnki.11-1235/f.2013.07.013

Lang, M.H., & Lundholm, R.J. (1996). Corporate disclosure policy and analyst behavior. *Accounting Review*, 71(4), 467–492. https://www.jstor.org/stable/248567

Lo, A.W., & MacKinlay, A.C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3(2), 175–205. https://doi.org/10.1093/rfs/3.2.175

Parsons, C.A., Sabbatucci, R., & Titman, S. (2020). Geographic lead-lag effects. *Review of Financial Studies*, 33(10), 4721–4770. https://doi.org/10.1093/rfs/hhz145

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003

Romer, P.M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2), 71–102. https://doi.org/10.1086/261725

Sias, R.W., & Starks, L.T. (1997). Return autocorrelation and institutional investors. *Journal of Financial Economics*, 46(1), 103–131. https://doi.org/10.1016/S0304-405X(97)00026-3

Van Nieuwerburgh, S., & Veldkamp, L. (2009). Information immobility and the home bias puzzle. *Journal of Finance*, 64(3), 1187–1215. https://doi.org/10.1111/j.1540-6261.2009.01462.x

Veldkamp, L.L. (2006). Information markets and the comovement of asset prices. *Review of Economic Studies*, 73(3), 823–845. https://doi.org/10.1111/j.1467-937X.2006.00397.x

Verrecchia, R.E. (1982). Information acquisition in a noisy rational expectations economy. *Econometrica*, 50(6), 1415–1430. https://doi.org/10.2307/1913389

Wang, C. (2014). Accounting standards harmonization and financial statement comparability: Evidence from transnational information transfer. *Journal of Accounting Research*, 52(4), 955–992. https://doi.org/10.1111/1475-679X.12055