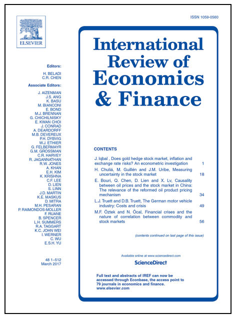
Journal Pre-proof




# Sell-Side Financial Analyst Social Network and Forecast Accuracy

Mengjia Li, Wenjie Ding, Hao Li, Qingwei Wang, Jason Zezhong Xiao

PII: S1059-0560(25)00664-1
DOI: https://doi.org/10.1016/j.iref.2025.104501
Reference: REVECO 104501

To appear in: *International Review of Economics and Finance*

Received Date: 18 June 2024
Revised Date: 17 July 2025
Accepted Date: 6 August 2025

# Sell-Side Financial Analyst Social Network and Forecast Accuracy[1]

Mengjia Li[3], Wenjie Ding[2], Hao Li[4], Qingwei Wang[5], and Jason Zezhong Xiao[6]

**Abstract**

Based on a unique dataset in China from 2012 to 2021, we find that sell-side financial analysts' social network improves analysts' forecast accuracy. Specifically, we find that analysts with a more central position in social networks based on corporate site visits generally have more face-to-face opportunities to learn from their peers, significantly improving their forecast performance. Such a social learning effect exists when more influential peers attend corporate site visits and when forecasted firms with higher information uncertainty.

**Keywords:** Sell-side analysts; Social network; Corporate site visits; Social learning; Forecast accuracy

**JEL codes:** G34; G41; O16; M41



---

[1] We thank Guiming Han, Fangfang Hou, Xinhe Huang, Zezeng Li, Oksana Pryshchepa, Theresa Spickers, Cong Wang and all participants at the 17th International Behavioural Finance Conference for constructive discussions. Wenjie Ding would like to thank the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 72203244) for funding this research. Every author has equally contributed to this paper. Errors and omissions remain the responsibility of the authors.

[2] Business School, Sun Yat-sen University. Email: dingwj23@mail.sysu.edu.cn. Corresponding Author. Sun Yat-sen University is the first affiliation of this paper.

[3] Faculty of Business Administration, University of Macau. Email: limengjia2017@hotmail.com.

[4] Cardiff Business School. Email: lih86@cardiff.ac.uk.

[5] Cardiff Business School. Email: wangq30@cardiff.ac.uk.

[6] Faculty of Business Administration, University of Macau. Email: jasonxiao@um.edu.mo.

**Sell-Side Financial Analyst Social Network and Forecast Accuracy**

## 1. Introduction

Sell-side financial analysts play a crucial role in capital markets, and their opinions have a significant impact on the valuation of assets (Bradshaw, 2004; Gleason and Lee, 2003; Jegadeesh et al. 2004; Stickel, 1992). Most of previous studies focus on the determinants of financial analysts' forecast performance by regarding individual analyst characteristics and information environment (Brown 1983; Brown and Rozeff 1979; Byard et al. 2011; Clement 1999; Hope 2003; Jacob et al. 1999; Lang and Lundholm 1996; Mikhail et al. 1997), while some researchers also indicate the significance of analysts' peer effects. That is, analysts are significantly concerned with the opinions of other analysts about the firms they cover (e.g. Graham, 1999; Zhao et al., 2014; Hou et al., 2018; Kumar et al., 2022).

There might be two explanations for analysts' peer effects in different scenarios. First, analysts' forecasts could be influenced by strategic herding behavior, where analysts pay more attention to peers' forecasts and recommendations for the same target firm (Graham, 1999; Trueman, 1994; Welch, 2000). Strategic herding behavior can arise from information cascades, which can also be called as "informational herding", where analysts infer information of the target firm from other analysts' earnings estimates (Bikhchandani et al., 1992). Alternatively, strategic herding behavior can arise from intentional and strategic behavior, where analysts are afraid to deviate from consensus for professional reasons (Hong et al., 2000). The ability to extract information from the current actions of others may be an important source of analyst expertise (Clement et al., 2011). Second, Kumar et al. (2022) proposed a quite new explanation for peer effects on different target firms. They argue that sell-side equity analysts engage in social learning to improve their forecast performance. Specifically, Kumar et al. (2022) indicate that an analyst's earnings forecast for a target firm is additionally influenced by the actions and opinions of peer analysts who follow the same firms in that analyst' following portfolio. For example, if the analyst follows a range of firms from firm $k_1$ to $k_{10}$, but only issues earnings forecasts for the firm $k_1$. That earnings forecast may be influenced by the actions and opinions of other peer analysts who also follow firm $k_1$ to $k_{10}$, or follow some of these firms, but not only those analysts who follow $k_{10}$. According to limited attention theory, analysts may pay more attention to other analysts' views on firms in their own following portfolio, but pay relatively less attention to similar information on other firms that are out of their portfolio. Therefore, the heterogeneity of analysts' following portfolios leads to the heterogeneity in analysts' information sets. Analysts can correct their bias by learning from peers. For example, if peer analysts are systematically optimistic (pessimistic) about other firms within the analysts' following portfolio, the analyst may learn from peers and update his views on the target firm, that is, issue a more pessimistic (optimistic) forecast to correct the perceived bias, thereby improving the accuracy of the forecast.

Social learning hypothesis, grounded from the widely acknowledged theory that social cognitive theory (Bandura, 1977), highlights the notion that people learn by observing and imitating others, particularly those held in an admired status, as a fundamental aspect of human learning. The finance and economics literature defines "social learning" as a process where individuals learn from others in a way that extends beyond pure informational herding (Ellison and Fudenberg, 1993; Kaustia and Rantala, 2015; Moretti, 2011), which suggests that analysts' forecast performance could also be improved by their social network. For example, Malmendier and Shanthikumar (2014) find that analysts may not learn from their own past mistakes, but they could learn from their peers. Do and Zhang (2020) demonstrate how the forecasting performance of existing analysts is influenced by the arrival or departure of star analysts. They argue that star analysts offer incumbent analysts role models and give them the opportunity to observe and learn (e.g., the star analyst's work ethic and way of interacting with clients and other members of the team). These tacit lessons are helpful in improving incumbents' overall performance.

Distinct from previous studies, our research provides more direct empirical evidence on the social learning hypothesis of analysts' peer effects. We attempt to recognize social learning beyond pure informational herding by examining the impact of private interactions among analysts on their overall forecast performance. We use a unique dataset of corporate site visits from the Shenzhen Stock Exchange (SZSE) to construct a social network of analysts based on face-to-face interactions among analysts during corporate site visits. Specifically, we first construct a social network of analysts based on their attendance of corporate site visits. We argue that analysts should socially connected with each other if they attend the same corporate site visits. Then, we calculate the eigenvector centrality of each analyst. The eigenvector centrality can fully account for indirect and direct social interactions. Finally, we examine the effect of analysts' eigenvector centrality on their overall forecast performance.

Our research is distinct from Cheng et al. (2016) in that we investigate the effect of analysts' social network rather than information acquisition from corporate site visits. Although we find similar results that corporate site visits contribute to improving analysts' earnings forecasts, the mechanism behind it might be different. Following Kumar et al. (2022), we recognize social learning but not informational herding through the improvement of analysts' overall forecast performance. That is, informational herding hypothesis could be the underlying mechanism only if analysts improve their forecast performance for the visited firm after corporate site visits, which is examined by Cheng et al. (2016). However, we do not impose any requirement that these analysts forecast visited firms after corporate site visits. In fact, most of analysts in our sample do not publish earnings estimates for visited firms. Therefore, the improvement of analysts' overall forecast performance for analysts with a more central position in the social network based on analysts' attendance of corporate site visits indeed signals the social learning explanation rather than informational herding hypothesis. In addition, our research is also distinct from Kumar et al. (2022) in that our proxy for peer analysts is more direct to test the social learning hypothesis. Kumar et al. (2022) define peer analysts as analysts who publish earnings estimates for same firms within analysts' following portfolios, while these analysts may not know each other personally, which means that analysts can only extract information from peers' public earnings estimates.

On the contrary, we define peer analysts if they attend the same corporate site visits. We believe that the face-to-face communications among analysts during corporate site visits can bring numerous new information, new knowledge, new opinions, and new sentiment, which should have more direct and stronger effects than peers' public earnings estimates.

In sum, we find that analysts' social network improves analysts' forecast accuracy. Specifically, analysts with higher eigenvector centrality in the social network based on corporate site visits generally provide more accurate earnings forecasts relative to other analysts. We conduct a battery of robustness tests to address potential empirical concerns. Our result is robust to the control of the firm' and analysts' characteristics that are commonly used in the previous studies and to the use of alternative measures of forecast accuracy. To alleviate the concern of sample selection bias that not all analysts have site visits before publishing earnings forecasts, we use the Heckman's Two-Step Selection method. Moreover, our model may suffer the endogeneity concern that analysts publish more accurate earnings forecasts will attend more corporate site visits. The unobservable omitted variables, for example, analysts' personality, may be also related to analysts' forecast performance and their position in the social network. Therefore, we employ the fixed effect model, instrumental variable, and subsamples to alleviate the concern of omitted variables and reverse causality. Following Han et al. (2018), we use the extreme weather as the instrumental variable to take out the endogenous effect because the extreme weather significantly affects the possibility of corporate site visits while seems not correlates with analysts' forecast performance. Following Chen et al. (2022) we restrict our sample to a subset of firm-quarter observations for which the reverse causation problem is less severe. Our conclusions do not alter after these robustness checks.

To further substantiate our main results relating to the social learning hypothesis, we examine two situations implied by the hypothesis through which social network improves analysts' forecast accuracy: influential peers and information uncertainty.

According to Centola (2010) and Aral and Walker (2012), influential peers have a significant effect on the diffusion of knowledge, ideas, and behaviors within social networks. They suggest that influential individuals not only possess more information but also have a greater ability to persuade others to adopt certain beliefs or practices. Moreover, Bonaccio and Dalal (2006) and Chen et al. (2022) demonstrate that individuals are more likely to seek advice from others and are more receptive to learning from the experiences and knowledge of influential peers in uncertain situations. Their studies highlight the role of uncertainty in driving individuals to actively seek and learn from others. Therefore, we follow these studies to argue that analysts should learn more from peers when more influential peers attended corporate site visits or when forecasted firms with higher information uncertainty.

This study contributes to the growing body of research on the social learning hypothesis in Finance. Distinct from the prior studies, this study provides a more direct proxy for analysts' peer effects. Unlike most of previous research define peer analysts if they issue earnings estimates for the same firms, this study objectively quantifies the peer effects based on face-to-face interactions with a unique dataset of corporate site visits in China. Although corporate site visits are common in the United States and Europe, firms usually do not report historical records of these visits. However, firms listed on the SZSE in China have been obligated to disclose information regarding site visits since 2012, creating a distinct prospect to scrutinize the direct interactions between analysts during these visits. The valuable dataset allows us to construct a powerful analysts' social network because we believe that analysts should have face-to-face communications and build strong relationships with each other if they attend the same corporate site visits. From psychological literature (Carr, 2011; Turkle, 2011), virtual and textual information, for example, public earnings estimates, often lacks the nuances and subtleties necessary for genuine peer effects to take place. Instead, face-to-face interactions have a stronger peer effect than virtual and textual information. Psychological literature emphasizes the unique qualities and depth of in-person communication, highlighting the limitations of digital media in fully capturing the richness of human interaction. To the best of our knowledge, this study is the first to construct the analysts' social network based on corporate site visits to measure analysts' peer effects.

Our results that analysts' face-to-face social network contributes to improving analysts' forecast performance has several implications for regulators, investors, firm managers, and financial analysts themselves. First, regulators can encourage or facilitate networking opportunities for financial analysts, such as organizing industry conferences or events where analysts can meet and interact face-to-face. By recognizing the value of social networks in improving forecast accuracy, regulators can promote a more collaborative and information-sharing environment within the financial industry. Second, investors can consider the social network of financial analysts as an additional factor when evaluating the quality of their forecasts. Investors may prioritize analysts who actively expand their face-to-face social networks, as these analysts are more likely to have access to diverse information sources and benefit from the exchange of insights and perspectives with influential peers. Third, firm managers can support and encourage financial analysts to engage in networking activities and build relationships with influential peers. Firms can facilitate opportunities for analysts to attend industry events, participate in professional organizations, or engage in cross-departmental collaboration within the organization. By fostering a culture of networking and knowledge sharing, firms can enhance the accuracy of their financial forecasts. Finally, financial analysts themselves can proactively expand their face-to-face social networks to improve their forecast accuracy. They can attend industry conferences, join professional organizations, and actively engage with influential peers in their field. By building strong relationships with knowledgeable and well-connected individuals, analysts can gain access to diverse information, receive feedback on their analyses, and benefit from the expertise and insights of others. Financial institutions, for example, brokers, can support and incentivize networking efforts by incorporating social network expansion as a performance metric or providing resources for analysts to attend relevant conferences and events. By recognizing the value of social networks, institutions can encourage analysts to invest time and effort into cultivating relationships that can enhance their forecast accuracy. Overall, the implication of knowing that expanding face-to-face social networks can improve analysts' forecast accuracy suggests the importance of collaboration, knowledge sharing, and relationship-building within the financial industry. By recognizing and leveraging the power of social networks, regulators, investors, firms' managers, and financial analysts can enhance the quality and reliability of financial forecasts.

The rest of the paper proceeds as follows. Section 2 covers data and variable definitions, section 3 discusses empirical results, section 4 presents some cross-sectional analyses, and section 5 concludes.

## 2. Data and variables

We obtain information on analysts' earnings forecasts and corporate site visits for all listed firms on Chinese SZSE market from fiscal years 2012-2021. We start our sample in 2012 because corporate site visits in earlier years are sparse in the CSMAR database. We include the analysts' latest published EPS forecasts for each fiscal year and no later than the fiscal year-end. Because we compare analysts' relative forecast performance for a particular firm within a year, we eliminate firm-years for which only one analyst provides a forecast. We remove analysts who did not attend any corporate site visits during the fiscal year because no network connection is constructed based on those analysts. Our final sample consists of 142,601 analyst-firm-year observation.

### 2.1 Social network and centrality measures

We construct analysts' social network based on their attendance at corporate site visits. To measure how well connected an analyst is in the social network based on corporate site visits, we follow Hirshleifer et al. (2021) and construct a network centrality degree, eigenvector centrality (EC), which is commonly used in graph theory to characterize the extent to which the prominence or importance of a node in the network. In the analysts' social network, the security analyst is selected as the node and with N = 1, . . . ,

n. The edge between analyst i and analyst j, denoted as $a_{ij}$, represents the connection between the two analysts based on corporate site visits.

Self-links or loops (a node transferring information to itself) are not allowed in the graph ($a_{ii} = 0$). The undirected ($a_{ij} = a_{ji}$) and weighted ties among analysts are reflected in the symmetric adjacency matrix $A = \{a_{ij}\}_{N\times N}$, that is:

$$A = \begin{Bmatrix} 0 & \cdots & a_{1i} & \cdots & a_{1n} \\ \vdots & \ddots & & & \vdots \\ a_{i1} & & 0 & & a_{in} \\ \vdots & & & \ddots & \vdots \\ a_{n1} & \cdots & a_{ni} & \cdots & 0 \end{Bmatrix}$$

where N is the number of analysts and $a_{ij}$ is the number of corporate site visits links between two analysts.

EC accounts for the transmission of signals along longer paths and walks (Bonacich, 1972; Borgatti, 2005). The EC of a node $i$ is the $i_{th}$ element of the principal right eigenvector of the adjacency matrix. The centrality of a node is also proportional to the average centrality scores of its direct neighbors. Therefore, a node will be more central if it is adjacent to nodes that are themselves highly central. The advantage of EC is that it fully allows for indirect and direct social interactions.



**2.2 Analysts' forecast accuracy**

Following Clement and Tse (2005), our baseline measure of an analyst i's forecast accuracy for firm k in year t is based on the absolute forecast error (AFE) of her forecast relative to those of others who follow firm k in year t. We first calculated AFE of analyst i for firm k in year t as:

$$AFE_{ikt} = |Forecasted\ EPS - Actual\ EPS|, \tag{1}$$

Then, we scale the difference between the maximum AFE of firm k and analyst i's AFE of firm k by the range of AFE for analysts following firm k in year t:

$$Accuracy_{ikt} = \frac{AFEmax_{kt} - AFE_{ikt}}{AFEmax_{kt} - AFEmin_{kt}},$$

*(2)*

In this way, $Accuracy_{ikt}$ increases with analyst i's own forecast performance. It measures the least accurate forecast (highest AFE) as 0 and the most accurate forecast (lowest AFE) as 1.

## 3. Empirical results

### 3.1 Baseline results

To assess whether analysts' forecast accuracy increases as a function of her eigenvector centrality in the social network, we estimate the following regression model:

$$Accuracy_{ikt} = \alpha + \beta_1 EC_{it} + \gamma Controls + Fixed\ Effects + \varepsilon_{ikt}.$$

*(3)*

Our controls for other determinants of analysts' relative accuracy include analysts' characteristics and firm characteristics. Following Clement and Tse (2005) and Hirshleifer et al. (2019), we control for analysts' characteristics including: analysts' forecast frequency (*ForFrequency*), forecast horizon (*ForHorizon*), firm-specific forecast experience (*FirmExperience*), general forecast experience (*GenExperience*), the number of firms (*FollowF*) and industries (*FollowI*) each analyst follows, the number of analysts covers a firm (*FollowA*), analysts' brokerage size (*BrokerSize*) and forecast accuracy in the prior year (*LagAccuracy*). Following Han et al. (2018) and Ding et al. (2021), we control for firm characteristics including: firm size (*size*), leverage (*LEV*), age (*Age*) and return on assets (*Roa*). Following Han et al. (2018) and Hirshleifer et al. (2019), we control for analyst-firm fixed effects and year fixed effects to account for unobserved analyst-firm and year heterogeneity.

We report descriptive statistics in Table 1. The average forecast accuracy in our sample is over 0.5, which suggests that analysts who attended corporate site visit have higher forecast accuracy above the average of peer analysts. However, other relative characteristics of analysts in our sample are all below 0.5, indicating that analysts who attend corporate site visits are generally less experienced (in both general and firm-specific experience), issue less frequent forecasts and more recent forecasts, follow fewer firms and industries, and in smaller brokerages.

[Insert Table 1 about here]

Column (1) and (2) in Table 2 indicate that, on average, the accuracy of forecast increases as a function of analysts' eigenvector centrality in the social network. In column (2), the coefficient on our key independent variable, *EC*, is 0.394 and is significant at the 1% level. This suggests that, on average, a one-unit increase in *EC* leads to a forecast that is 0.394 units more accurate relative to others. This is an economically meaningful effect. This result supports our hypothesis that analysts who are more central in the social network provide more accurate earnings forecasts relative to others.

[Insert Table 2 about here]

## 3.2 Robustness tests

### 3.2.1 Alternative measures of forecasts accuracy

In our baseline regression, we follow Clement and Tse (2005) to define forecast accuracy as expressed in Equation (1) and (2). In this section, we employ two alternative measures of forecast accuracy. First, following Han et al. (2018), we replace the AFE in $Accuracy_{ikt}$ by the AFE scaled by share price of firm k in two days before the forecast, other calculations are the same as $Accuracy_{ikt}$. Second, following Kumar et al. (2022), we measure forecast accuracy as the average AFE for analysts who follow firm *k* in year t minus the AFE of analyst *i* following firm *k* in year *t*, with this difference scaled by the average of AFE for analysts following firm *k* in year *t*, expressed as:

$$Accuracy_{ikt} = \frac{AFEmean_{kt} - AFE_{ikt}}{AFEmean_{kt}}, \tag{4}$$

Results in Table 3 show that our results are robust to all three alternative measures of forecast accuracy. For brevity, we report only the coefficient estimates for the main variables of interest. All regression results in Table 3 are consistent with our main hypothesis that more analysts with higher eigenvector centrality provide more accurate earnings forecasts relative to their peers.

[Insert Table 3 about here]

### 3.2.2 Heckman two-staged procedure

Not all analysts have site visits when making earnings forecasts. We cannot estimate analysts' eigenvector centrality score of the social network based on corporate site visits if they have no corporate site visits. This raises the question whether the differences in the characteristics of these two groups of analysts drive our results. To alleviate the sample self-selection concern, we use the Heckman two-step method. We follow Cheng et al. (2016, 2019) and use the following regression to estimate the probability of analysts attending corporate site visits and obtain the Inverse Mills Ratio (IMR):

$$P\,r\left(EC_{treat_{it}}\right) = \alpha + \beta_1 Num\_Firms_{kt} + \beta_2 \Delta GDP_{kt} + \gamma Controls + \varepsilon_{ikt}, \qquad (5)$$

where $EC\_treat_{it}$ is an indicator variable coded 1 if the analyst $i$ has at least one corporate site visit to measure her eigenvector centrality in year $t$, and 0 otherwise. For determinants, we add two instruments that the information related to firm headquarters' city, including the number of listed firms ($Num\_Firms_{kt}$) and GDP growth ($\Delta GDP_{kt}$). Following Jiang and Yuan (2018), Cheng et al. (2019) and Chen et al. (2022), these two variables are exclusion restrictions. We add these two variables in the first stage because they are expected to correlate with analysts' eigenvector centrality score of the social network based on corporate site visits, but they are not directly related to analysts' forecast accuracy. For example, more listed firms in the firm headquarters' location can attract more site visits, because analysts prefer to visit cities where they can visit multiple firms in one trip to save time and expenses, while it is not directly related to analysts' forecast accuracy. Similarly, the changes in cities' GDP where firm headquarters is located attract more site visits to explore reasons behind, while it is not directly related to analysts' forecast accuracy.

Table 4 reports the results. Colum (1) presents the determinant analysis. As we expected, $Num\_Firms_{kt}$ and $\Delta GDP_{kt}$ are both significantly related to the probability of analysts attending corporate site visits. In the second stage, we test the effect of analysts' eigenvector centrality score on forecast accuracy by including the IMR estimated from the first step. Column (2) of Table 4 shows that, similar to the baseline results reported in Table 4, the coefficient of $EC_{it}$ is positive and statistically significant at the 1% level. These results show that our baseline findings are robust when using the Heckman two-staged procedure to adjust for the self-selection bias.

[Insert Table 4 about here]

### 3.2.3 Instrumental variable

One may concern that the omitted variables in the baseline models that are related to analysts' eigenvector centrality of the social network also affect forecast accuracy. Although we include analyst-firm and year fixed effects to alleviate concerns that forecast accuracy is driven by time and analyst-firm invariant unobservable variables, there may be other omitted variables that lead to reverse causality. For instance, analysts with higher level of professional skills attend more corporate site visits and provide more accurate forecasts. To alleviate this endogeneity concern, we employ the instrumental variable and two-stage least square method.

We use the instrumental variable approach to identify the causal relationship between analysts' eigenvector centrality and forecast accuracy. Following Han et al. (2018), we use an exogenous variable, extreme weather conditions (*ExtrmWeather*) in the city of the firm headquarter, as an instrument for corporate site visits. Weather affects the probability of corporate site visits, which affects analysts' eigenvector centrality score of the social network based on corporate site visits, as it is more difficult to travel to places during extreme weather. However, weather is unlikely to affect analysts' forecast accuracy. Thus, we expect extreme weather to represent a valid IV estimation of analysts' eigenvector centrality. We define a day as an extreme weather day (*ExtrmDay* = 1) if the lowest temperature falls below -10°C or if the highest temperature reaches above 37°C. *ExtrmWeather* is defined as the percentage of days in year t with extreme weather conditions in the city where the firm's headquarters is located:

$$ExtrmWeather_{kt} = \frac{\sum ExtrmDay_{kt}}{TotalDays_t}. \tag{6}$$

We use the quintile rank of *ExtrmWeather* as the instrumental variable. Table 5 presents the results. Column (1) of Table 5 reports the results of the first-stage regressions where the dependent variable is analysts' eigenvector centrality score, and the explanatory variables include the instrument and the same set of control variables as in Table 2. For brevity, we report only the coefficient estimates for the main variables of interest. Consistent with the rationale behind the instrument, *ExtrmWeather* is positively and significantly (at the 1% level) correlated to analysts' eigenvector centrality, suggesting that our instrument is valid. The reported F-statistics are large, the p-value of the Cragg-Donald's Wald F weak-instrument test statistic is 0.000, both rejecting the null hypothesis that the instrument is weak (Cragg and Donald, 1993; Stock and Yogo, 2005).

Column (2) of Table 5 reports the results for the second-stage regressions with analysts' forecast accuracy as dependent variable. The variable of interest is the variable with the predicted values from the regression in the first-stage regressions. The results are consistent with the baseline regressions and support our main hypothesis. Those results imply that our key result is unlikely due to the endogeneity of the analysts' social network.

[Insert Table 5 about here]

### 3.2.4 Reverse causality

Whereas all our identification attempts so far point to a causal effect of the eigenvector centrality of analysts on their forecast accuracy, a plausible alternative interpretation of our main results is that analysts who are more accurate in their earnings forecast attend more corporate site visits, resulting in the positive relation between the centrality of analysts and forecast accuracy. This alternative interpretation indicates the direction of causality could be the other way around. To gain insights about whether our findings are driven by reverse causality, we follow Chen et al. (2022) to restrict our sample to a subset of firm-quarter observations for which the reverse causation problem is less severe. More specifically, we re-examine the effects of analysts' eigenvector centrality after excluding, respectively, the top 10% and 25% accurate analysts. We report the results in Table 6. We find that the eigenvector centrality of analysts based on the social network of corporate site visits continues to be economically and statistically significant in all model specifications. These findings provide further assurance that the effect of CSV does not appear to arise from reverse causation.

[Insert Table 6 about here]

## 4. Cross-sectional analysis of social learning hypothesis

Our previous results are consistent with the social learning hypothesis that sell-side analysts learn from their peers to improve forecast accuracy. In this section, we conduct a series of cross-sectional analyses to further validate the social learning channel.

### 4.1 Influential peers

Psychological literature indicates the significant peers effect of influential peers. For example, Centola (2010) conducted a large-scale online social network experiment to examine how different types of influence shape behavior adoption. The results showed that participants were more likely to adopt a behavior when they were exposed to influential peers who had already adopted that behavior. Aral and Walker (2012) conducted a study to identify influential and susceptible individuals within social networks and examine their impact on behavior adoption. The research combined large-scale data from an online social network with a randomized experiment. The findings revealed that influential peers had a greater effect on behavior adoption compared to non-influential peers, supporting the argument that people learn more from influential peers. Therefore, we identify influential peers in corporate site visits and expect to see influential peers significantly affect analysts to improve forecast accuracy under the social learning hypothesis.

Following Chen et al. (2022), we recognize influential analysts as analysts with more expertise. Therefore, we identify influential peers if their affiliations are top 10 brokers, if they are star analysts, if they have a PhD degree, or if they are experienced analysts. Based on the sample median of these proxies we run split sample regressions, and the results are shown in Table 7.

The results in Table 7 are in line with our expectations. The coefficients of *EC* on *Accuracy* are all significantly positive in subsamples with more influential analysts, and all insignificant in subsamples with less influential analysts. Hence, analysts with a higher eigenvector centrality in the social network based on corporate site visits forecast more accurate than others if the percentage of analysts from top 10 brokers is higher, the percentage of star analysts is higher, the percentage of analysts with a PhD degree is higher, or the percentage of experienced analysts is higher in corporate site visits. It shows that analysts learn more from influential peers, that is, peers with more expertise.

[Insert Table 7 about here]

**4.2 Information uncertainty**

Previous literature indicates that individuals tend to look to others for cues on how to behave in uncertain situations. Moreover, Bandura (1977) suggests that efficacy expectation can vary because of the level of difficulty of the task. Previous literature also confirmed that learning from others provides diverse perspectives that enhances individual learning outcomes when individuals face challenging tasks (Bonaccio and Dalal, 2006). Therefore, we argue that analysts have more motivations to learn from peers when forecasting earnings is more difficult. We expect that analysts should learn more from peers when forecasted firms with higher information uncertainty.

We measure the difficulty level of forecasting a firm by the its information uncertainty, which is higher if the firm has larger volatility of daily stock returns, larger volatility of adjusted ROA, or it is a larger firm or younger firm. Based on the sample median of these proxies we run split sample regressions, and the results are shown in Table 8.

The results in Table 8 align with our expectations. The coefficients of *EC* on *Accuracy* are all significantly positive in subsamples with higher information uncertainty, and all insignificant in subsamples with lower information uncertainty. It suggests that analysts with a higher eigenvector centrality in the social network based on corporate site visits forecast more accurate than others if the firm has larger volatility of daily stock returns, larger volatility of adjusted ROA, or is a larger firm or a younger firm. It shows that analysts learn more from peers if the firm is more difficult to forecast.

[Insert Table 8 about here]

## 5 Conclusion

This paper employs a unique dataset of site visits disclosed in China to examine the role of sell-side analysts' social network on analysts' forecast accuracy. We find that analysts with higher eigenvector centrality of the social network based on corporate site visits provide more accurate earnings forecasts. The results are robust to alternative measures of forecast accuracy, the change of accuracy, Heckman two-staged tests, instrumental variable, and subsample analysis that aim at addressing reverse causality.

We then conduct a battery of tests to uncover the underlying mechanism for the relationship between centrality and forecast accuracy. We find supporting evidence for the social learning mechanism. We find that the effect of social network on forecast accuracy exists when: 1) there are more influential peers in corporate site visits; 2) forecasted firms have higher information uncertainty.

Our study highlights the positive effect of direct interactions between analysts. We show that analysts with higher centrality of social network provide more accurate earnings forecasts, which has important implications for investors, managers and regulators.

**Table 1: Summary statistics**

| | N | Mean | SD | Min | p5 | p25 | Median | p75 | p95 | Max |
|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy | 142601 | 0.638 | 0.330 | 0.000 | 0.000 | 0.390 | 0.741 | 0.929 | 1.000 | 1.000 |
| EC | 142601 | 0.005 | 0.015 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.033 | 0.167 |
| ForFrequency | 142601 | 0.307 | 0.322 | 0.000 | 0.000 | 0.032 | 0.200 | 0.500 | 1.000 | 1.000 |
| ForHorizon | 142601 | 0.374 | 0.338 | 0.000 | 0.000 | 0.092 | 0.255 | 0.668 | 1.000 | 1.000 |
| FirmExperience | 142601 | 0.229 | 0.333 | 0.000 | 0.000 | 0.000 | 0.000 | 0.333 | 1.000 | 1.000 |
| GenExperience | 142601 | 0.263 | 0.277 | 0.000 | 0.000 | 0.077 | 0.167 | 0.375 | 1.000 | 1.000 |
| FollowF | 142601 | 0.326 | 0.274 | 0.000 | 0.000 | 0.118 | 0.250 | 0.460 | 1.000 | 1.000 |
| FollowI | 142601 | 0.322 | 0.273 | 0.000 | 0.000 | 0.115 | 0.250 | 0.462 | 1.000 | 1.000 |
| FollowA | 142601 | 49.335 | 36.464 | 2.000 | 8.000 | 22.000 | 41.000 | 68.000 | 119.000 | 290.000 |
| BrokerSize | 142601 | 0.479 | 0.311 | 0.000 | 0.000 | 0.239 | 0.450 | 0.698 | 1.000 | 1.000 |
| LagAccuracy | 142601 | 0.721 | 0.275 | 0.000 | 0.071 | 0.579 | 0.813 | 0.937 | 1.000 | 1.000 |
| Size | 142601 | 23.034 | 1.602 | 19.321 | 21.012 | 21.923 | 22.740 | 23.783 | 26.105 | 31.191 |
| LEV | 142601 | 0.430 | 0.199 | 0.009 | 0.127 | 0.272 | 0.420 | 0.576 | 0.771 | 2.579 |
| Age | 142601 | 17.359 | 5.784 | 3.000 | 8.000 | 13.000 | 17.000 | 21.000 | 27.000 | 63.000 |
| Roa | 142601 | 0.067 | 0.062 | -3.911 | 0.006 | 0.033 | 0.060 | 0.095 | 0.165 | 0.590 |

This table presents descriptive statistics on variables. *Accuracy* is analyst i's forecasts' accuracy for firm k in year t relative to other analysts following firm k in year t. *Centrality* is the eigenvector centrality based on the network of corporate site visits for each analyst i in year t. *ForFrequency* is analyst i's forecast frequency for firm k in year t relative to other analysts following firm k in year t. *ForHorizon* is the time from the forecast date to the end of the fiscal period for analyst i following firm k in year t relative to other analysts following firm k in year t. *FirmExperience* is the number of years of firm specific experience for analyst i following firm k in year t relative to other analysts following firm k in year t. *GenExperience* is the number of years of experience for analyst i following firm k in year t relative to other analysts following firm k in year t. *FollowF* is the number of companies followed by analyst i following firm k in year t relative to other analysts following firm k in year t. *FollowI* is the number of industries followed by analyst i following firm k in year t relative to other analysts following firm k in year t. *FollowA* is the number of analysts who cover firm k in year t. BrokerSize is the number of analysts employed by the brokerage employing analyst i following firm k in year t relative to other analysts following firm k in year t. *LagAccuracy* is analyst i's forecasts' accuracy for firm k in year t-1 relative to other analysts following firm k in year t-1. *Size* is the natural log of firm k's total assets at the end of the fiscal year t. *LEV* is the debt-to-assets ratio of firm k at the end of the fiscal year t. *Age* is the number of years from firm k's listed year to the year t. *Roa* is the income before extraordinary items deflated by total assets of firm k at the end of the fiscal year t.

**Table 2: Baseline regression results**

| *Dependent Variable: Accuracy* | | |
|---|---|---|
| | *(1)* | *(2)* |
| EC | 0.934*** | 0.342** |
| | (0.220) | (0.151) |
| ForFrequency | | -0.318*** |
| | | (0.006) |
| ForHorizon | | -0.003 |
| | | (0.005) |
| FirmExperience | | -0.410*** |
| | | (0.005) |
| GenExperience | | -0.005 |
| | | (0.009) |
| FollowF | | -0.038** |
| | | (0.018) |
| FollowI | | -0.019 |
| | | (0.012) |
| FollowA | | -0.006 |
| | | (0.010) |
| BrokerSize | | 0.001*** |
| | | (0.000) |
| LagAccuracy | | -0.007 |
| | | (0.009) |
| Size | | -0.006 |
| | | (0.008) |
| LEV | | -0.032 |
| | | (0.022) |
| Age | | -0.011*** |
| | | (0.002) |
| Roa | | 0.252*** |
| | | (0.041) |
| _cons | 0.900*** | 1.345*** |
| | (0.012) | (0.181) |
| Analyst-Firm FE | Yes | Yes |
| Year FE | Yes | Yes |
| Observations | 142601 | 142601 |
| Adjusted $R^2$ | 0.039 | 0.304 |

This table reports the effect of analyst i's centrality on her forecast accuracy. Column (1) reports the result of univariate regression. Column (2) reports the result of Equation (3): $Accuracy_{ikt} = \alpha + \beta_1 EC_{it} + \gamma Controls + Fixed\ Effects + \varepsilon_{ikt}$. Variable definitions can be found in Table A1 of the Appendix. The standard errors in brackets are clustered at the analyst level. ***, **, * indicate the coefficients are significant at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed statistical tests.

**Table 3: Alternative measures of forecast accuracy**

| | *Accuracy2* | *Accuracy3* |
|---|---|---|
| | *(1)* | *(2)* |
| EC | 0.284* | 1.034** |
| | (0.156) | (0.433) |
| | | |
| Controls | Yes | Yes |
| Analyst-Firm FE | Yes | Yes |
| Year FE | Yes | Yes |
| | | |
| Observations | 118018 | 142601 |
| Adjusted $R^2$ | 0.295 | 0.276 |

This table reports the robustness test results when using alternative measures of analysts' forecast accuracy. Variable definitions can be found in Table A1 of the Appendix. The standard errors in brackets are clustered at the analyst level. ***, **, * indicate the coefficients are significant at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed statistical tests.

**Table 4: Heckman two-staged procedure**

| | *EC_treat* | *Accuracy* |
|---|---|---|
| | *(1)* | *(2)* |
| Num_Firms | 0.046*** | |
| | (0.003) | |
| ΔGDP | -0.959*** | |
| | (0.109) | |
| EC | | 0.330** |
| | | (0.149) |
| IMR | | 0.173*** |
| | | (0.039) |
| | | |
| Controls | Yes | Yes |
| Analyst-Firm FE | No | Yes |
| Year FE | No | Yes |
| | | |
| Observations | 179453 | 131946 |
| Pseudo /Adjusted $R^2$ | 0.061 | 0.310 |

This table reports the results of Heckman two-staged procedure. Column (1) reports the determinant analysis of the probability of analysts attending corporate site visits. It presents the logistic regression results for Equation (7): $Pr(EC\_treat_{it}) = \alpha + \beta_1 Num\_Firms_{kt} + \beta_2 \Delta GDP_{kt} + \gamma Controls + \varepsilon_{ikt}$. Column (2) reports the effect of analysts' eigenvector centrality on forecast accurcy by including the inverse Mill's ratio (IMR). Variable definitions can be found in Table A1 of the Appendix. The standard errors in brackets are clustered at the firm level. ***, **, * indicate the coefficients are significant at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed statistical tests.

**Table 5: Instrumental variable**

| | First stage | Second stage |
|---|---|---|
| | *EC* | *Accuracy* |
| | *(1)* | *(2)* |
| ExtrmWeather | 0.022*** | |
| | (0.005) | |
| EC (Fitted) | | 19.174*** |
| | | (5.673) |
| Controls | Yes | Yes |
| Analyst-Firm FE | Yes | Yes |
| Year FE | Yes | Yes |
| Observations | 130267 | 130267 |
| Adjusted $R^2$ | 0.027 | 0.302 |
| F-statistic | 586.66*** | |
| Cragg-Donald (CD) Wald F-statistic | 26.596 | |
| Stock and Yogo (2005) weak ID test critical value | 16.38 | |

This table reports the results of instrumental variables method based on two-stage least squares (2SLS) panel regressions. Column (1) presents the first-stage regression results in which the dependent variable is *EC*. Column (2) reports the second-stage regression results. Variable definitions can be found in Table A1 of the Appendix. The standard errors in brackets are clustered at the firm level. ***, **, * indicate the coefficients are significant at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed statistical tests.

## Table 6: Excluding top analysts

| | *Dependent variable: Accuracy* | |
|---|---|---|
| | *Excluding largest 10%* | *Excluding largest 25%* |
| | *(1)* | *(2)* |
| EC | 0.401** | 0.449*** |
| | (0.158) | (0.168) |
| Controls | Yes | Yes |
| Analyst-Firm FE | Yes | Yes |
| Year FE | Yes | Yes |
| Observations | 125856 | 106761 |
| Adjusted $R^2$ | 0.313 | 0.283 |

This table reports the regression results by excluding the most accurate analysts. The top 10% analysts are measured as analysts who issue the top 10% accurate earnings forecasts. The top 25% analysts are measured as analysts who issue the top 25% accurate earnings forecasts. Variable definitions can be found in Table A1 of the Appendix. The standard errors in brackets are clustered at the firm level. ***, **, * indicate the coefficients are significant at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed statistical tests.

## Table 7: Influential peers

| | Dependent variable: Accuracy | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *Top_10* | | *Star* | | *PhD* | | *Experienced* | |
| | *High* | *Low* | *High* | *Low* | *High* | *Low* | *High* | *Low* |
| | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* | *(7)* | *(8)* |
| EC | 0.530** | -0.008 | 0.682** | 0.263 | 0.473* | 0.447 | 0.583** | 0.143 |
| | (0.223) | (0.284) | (0.292) | (0.245) | (0.254) | (0.281) | (0.294) | (0.234) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Analyst-Firm FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 69418 | 73183 | 53666 | 88935 | 68600 | 74001 | 62571 | 80030 |
| Adjusted$R^2$ | 0.371 | 0.273 | 0.372 | 0.285 | 0.328 | 0.287 | 0.375 | 0.296 |

This table presents the results of cross-sectional analyses for the social learning hypothesis by considering influential peers. Column (1) and (2) use the percentage of analysts from top 10 brokers to measure influential peers and run the split sample regressions based on its sample median. Column (3) and (4) use the percentage of star analysts to measure influential peers and run the split sample regressions based on its sample median. Column (5) and (6) use the percentage of analysts with a PhD degree to measure influential peers and run the split sample regressions based on its sample median. Column (7) and (8) use the percentage of analysts with more than 5 years of forecast experience to measure influential peers and run the split sample regressions based on its sample median. Variable definitions can be found in Table A1 of the Appendix. The standard errors in brackets are clustered at the firm level. ***, **, * indicate the coefficients are significant at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed statistical tests.

## Table 8: Information uncertainty

| | *Dependent variable: Accuracy* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *Firm risk1: Daily_stock_return* | | *Firm risk2: Adjusted_ROA* | | *Large_firms* | | *Young_firms* | |
| | *High* | *Low* | *High* | *Low* | *High* | *Low* | *High* | *Low* |
| | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* | *(7)* | *(8)* |
| EC | 0.462** | -0.223 | 0.499** | 0.155 | 0.566** | 0.214 | 0.397** | 0.249 |
| | (0.183) | (0.441) | (0.239) | (0.248) | (0.271) | (0.188) | (0.190) | (0.261) |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Analyst-Firm FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 75044 | 67557 | 78157 | 64444 | 71293 | 71308 | 74203 | 68398 |
| Adjusted$R^2$ | 0.325 | 0.268 | 0.329 | 0.268 | 0.264 | 0.356 | 0.350 | 0.270 |

This table presents the results of cross-sectional analyses for the social learning hypothesis by considering information uncertainty. Column (1) and (2) use the natural log of standard deviation of firm k's daily stock returns to measure firm risk and run the split sample regressions based on its sample median. Column (3) and (4) use the standard deviation of firm k's adjusted ROA to measure firm risk and run the split sample regressions based on its sample median. Column (5) and (6) use the size of firm k to measure firm k's information uncertainty and run the split sample regressions based on its sample median. Column (7) and (8) use the age of firm k to measure firm k's information uncertainty and run the split sample regressions based on its sample median. Variable definitions can be found in Table A1 of the Appendix. The standard errors in brackets are clustered at the firm level. ***, **, * indicate the coefficients are significant at the 0.01, 0.05, and 0.10 levels, respectively, based on two-tailed statistical tests.

## Appendix

### Table A1: Definitions of variables

| *Variables* | *Definitions* |
|---|---|
| ***Panel A: Variable definitions for the baseline regressions*** | |
| $Accuracy_{ikt}$ | Following Clement and Tse (2005), this measure is calculated as the maximum absolute forecast error for analysts who follow firm k in year t minus the absolute forecast error of analyst i following firm k in year t, with this difference scaled by the range of absolute forecast errors for analysts following firm k in year t. |
| $EC_{it}$ | The eigenvector centrality based on the network of corporate site visits for each analyst i in year t. |
| $ForFrequency_{ikt}$ | It is a measure of analyst i's forecast frequency for firm k, calculated as the number of firm k forecasts made by analyst i following firm k in year t minus the minimum number of firm-j forecasts for analysts following firm k in year t, with this difference scaled by the range in the number of firm-j forecasts issued by analysts following firm k in year t. |
| $ForHorizon_{ikt}$ | It is a measure of the time from the forecast date to the end of the fiscal period, calculated as the forecast horizon (days from the forecast date to the fiscal year-end) for analyst i following firm k in year t minus the minimum forecast horizon for analysts who follow firm k in year t, with this difference scaled by the range of forecast horizons for analysts following firm k in year t. |
| $FirmExperience_{ikt}$ | It is a measure of analyst i's firm specific experience, calculated as the number of years of firm specific experience for analyst i following firm k in year t minus the minimum number of years of firm specific experience for analysts following firm k in year t, with this difference scaled by the range of years of firm specific experience for analysts following firm k in year t. |
| $GenExperience_{ikt}$ | It is a measure of analyst i's general experience, calculated as the number of years of experience for analyst i following firm k in year t minus the minimum number of years of experience for analysts following firm k in year t, with this difference scaled by the range of years of experience for analysts following firm k in year t. |
| $FollowF_{ikt}$ | It is a measure of the number of companies analyst i follows in year t, calculated as the number of companies followed by analyst i following firm k in year t minus the minimum number of companies followed by analysts who follow firm k in year t, with this difference scaled by the range in the number of companies followed by analysts following firm k in year t. |
| $FollowI_{ikt}$ | It is a measure of the number of industries analyst i follows in year t, calculated as the number of industries followed by analyst i following firm k in year t minus the minimum number of industries followed by analysts who follow firm k in year t, with this difference scaled by the range in the number of industries followed by analysts following firm k in year t. The industry classification is based on the CSRC 2012 two-digit industry code. |
| $FollowA_{kt}$ | The number of analysts who cover firm k in year t. |
| $BrokerSize_{ikt}$ | It is a measure of the analyst's brokerage size, calculated as the number of analysts employed by the brokerage employing analyst i following firm k in year t minus the minimum number of analysts employed by brokerages for analysts following firm k in year t, with this difference scaled by the range of brokerage size for analysts following firm k in year t. |
| $LagAccuracy_{ikt}$ | It is a measure of analyst i's prior year forecast accuracy for firm k, calculated as the maximum Accuracy for analysts who follow firm k in year t-1 minus the Accuracy for analyst i following |

| *Variables* | *Definitions* |
|---|---|
| | firm k in year t-1, with this difference scaled by the range of Accuracy for analysts following firm k in year t-1. This measure is replaced by the median of analysts' prior year forecast accuracy for firm k if it has missing value. |
| $Size_{kt}$ | Natural logarithm of firm k's total assets at the end of the fiscal year t. |
| $LEV_{kt}$ | Debt-to-assets ratio of firm k at the end of the fiscal year t. |
| $Age_{kt}$ | The number of years from firm k's listed year to the year t. |
| $Roa_{kt}$ | Income before extraordinary items deflated by total assets of firm k at the end of the fiscal year t. |
| ***Panel B: Variable definitions for alternative proxies for forecasts' accuracy*** | |
| $Accuracy2_{ikt}$ | This measure replaces the absolute forecast error in $Accuracy_{ikt}$ by the absolute forecast error scaled by share price of firm k in two days before the forecast, other calculations are the same as $Accuracy_{ikt}$. |
| $Accuracy3_{ikt}$ | Following Kumar et al. (2022), this measure is calculated as the average absolute forecast error for analysts who follow firm k in year t minus the absolute forecast error of analyst i following firm k in year t, with this difference scaled by the average of absolute forecast errors for analysts following firm k in year t. |
| ***Panel D: Variable definitions for Heckman two-staged procedure*** | |
| $EC\_treat_{it}$ | An indicator variable coded 1 if the analyst i has at least one corporate site visit to measure her eigenvector centrality in year t, and 0 otherwise. |
| $Num\_Firms_{kt}$ | The number of listed firms in the province where the firm k's headquarters is, scaled by 100, in year t. |
| $\Delta GDP_{kt}$ | The growth of GDP of the city where the firm k's headquarters is, calculated as the city's GDP in year t divided by the GDP in year t-1, minus 1. |
| ***Panel E: Variable definitions for instrumental variable*** | |
| *ExtrmWeather* | First, we identify days with extreme weather conditions for each city where firm k's headquarters is located, if the lowest temperature falls below -10℃ or if the highest temperature reaches above 37℃. Second, we calculate the percentage of days in a year t with extreme weather conditions for each city. Finally, we use the quintile rank of the percentage of days scaled by 100 as the instrumental variable. |
| ***Panel F: Variable definitions for influential peers*** | |
| *Top_10* | The percentage of analysts from top 10 brokers in year t. |
| *Star* | The percentage of star analysts in year t. |
| *PhD* | The percentage of analysts with a PhD degree in year t. |
| *Experienced* | The percentage of analysts with more than 5 years of forecast experience in year t. |
| ***Panel G: Variable definitions for firm characteristics*** | |
| *Daily_stock_return* | We use the natural logarithm of the standard deviation of daily stock return to measure firm risk (FR1). $FR1_{kt} = \ln(\sqrt{\frac{1}{T}\sum_{\mathrm{d}=1}^{T}(r_{kdt} - \frac{1}{T}\sum_{\mathrm{d}=1}^{T} r_{kdt})^2})$ where $FR1_{kt}$ is the daily stock return of firm k on day d in year t. T is the number of total days in year t. |

| Variables | Definitions |
|---|---|
| *Adjusted_ROA* | We also follow John et al. (2008) to use the standard deviation of adjusted ROA to represent firm risk (FR2). $Adj\_ROA_{kt} = \frac{EBIT_{kt}}{ASSET_{kt}} - \frac{1}{X}\sum_{x=1}^{X}\frac{EBIT_{kt}}{ASSET_{kt}}$, where $Adj\_ROA_{kt}$ is the ROA of the firm k in year t minus annual industry average. In addition, the standard deviation of industry-adjusted ROA ($Adj\_ROA_{kt}$) is calculated separately on a rolling basis using every five years (from year t-4 to t) as an observation period. The firm risk is calculated by $FR2_{kt} = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(Adj\_ROA_{kt} - \frac{1}{T}\sum_{t=1}^{T}Adj\_ROA_{kt})^2}\,|T = 5.$ |
| *Large_firms* | An indicator variable coded 1 if the firm k's size is larger than the median of the full sample, and 0 otherwise. |
| *Young_firms* | An indicator variable coded 1 if the firm k's age is younger than the median of the full sample, and 0 otherwise. |

**References**

Aral, S. and Walker, D. 2012. Identifying influential and susceptible members of social networks. *Science* 337(6092), pp. 337-341.

Bandura, A. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84(2), pp. 191-215.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100 (5), pp. 992–1026.

Bonaccio, S. and Dalal, R. S. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101(2), pp. 127-151.

Bonacich, P. 1972. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* 2, pp. 113–20

Borgatti, S. P. 2005. Centrality and network flow. *Social Networks* 27, pp. 55 – 71.

Bradshaw, M. T. 2004. How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review* 79 (1), pp. 25–50.

Brown, L. D. 1983. Accounting changes and the accuracy of analyst's earnings forecasts. *Journal of Accounting Research* 21 (2), pp. 432–443.

Brown, L. D., and M. S. Rozeff. 1979. Adaptive expectations, time-series models, and analyst forecast revision. *Journal of Accounting Research* 17 (2), pp. 341–351.

Byard, D., Y. Li, and Y. Yu. 2011. The effect of mandatory IFRS adoption on financial analysts' information environment. *Journal of Accounting Research* 49 (1), pp. 69–96.

Carr, N. 2011. *The shallows: What the Internet is doing to our brains.* W. W. Norton & Company.

Centola, D. 2010. The spread of behavior in an online social network experiment. *Science* 329(5996), pp. 1194-1197.

Chen, J., Su, X., Tian, X., Xu, B. 2022. Does customer-base structure influence managerial risk-taking incentives? *Journal of Financial Economics* 143(1), pp. 462-483.

Chen, X., Cheng, C. S. A., Xie, J., and Yang, H. 2022. Private communication and management forecasts: Evidence from corporate site visits. *Corporate Governance: An International Review* 30(4), pp. 482– 497.

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

Cheng, Q., Du, F., Wang, Y., Wang, X. 2019. Do corporate site visits impact stock prices? *Contemporary Accounting Research* 36(1), pp. 359-388.

Clement, M. B. 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics* 27 (3), pp. 285–303, .

Clement, M. B. and Tse, S. Y. 2005. Financial analyst characteristics and herding behavior in forecasting. *Journal of Finance* 60(1), pp. 307-341.

Clement, M. B., Hales, J., Xue, Y. 2011. Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings. *Journal of Accounting and Economics* 51 (3), pp. 279-299.

Cragg, J. G. and Donald, S. G. 1993. Testing identifiability and specification in instrumental variable models. *Economic Theory* 9, pp. 222–240.

Das, S., Levine, C. B., Sivaramakrishnan, K. 1998. Earnings predictability and bias in analysts' earnings forecasts. *The Accounting Review* 73, pp. 277–294.

Ding, X., Tan, W., and Kang, Y. 2021. The spillover effect of regulatory penalties on management and analysts' earnings forecasts: Empirical evidence based on directors networks in China. *International Review of Economics & Finance* 76, pp. 502-515.

Do, T.T. and Zhang, H. 2020. Peer Effects among Financial Analysts. *Contemporary Accounting Research* 37, pp. 358-391.

Ellison G. and Fudenberg D. 1993. Rules of thumb for social learning. *Journal of Political Economy* 101 (4), pp. 612 – 643.

Glanz, K., Rimer, B. K., and Viswanath, K. V. 2015. Theory, research, and practice in health behavior. In Glanz, K., Rimer, B. K., and Viswanath, K. V. (Eds.), *Health behavior: Theory, research, and practice* (pp. 23–41).

Gleason, C. A., and C. M. C. Lee. 2003. Analyst forecast revisions and market price discovery. *The Accounting Review* 78 (1), pp. 193–225.

Graham, J.R., 1999. Herding among investment newsletters: theory and evidence. *Journal of Finance* 54, pp. 237–268.

Han, B., Kong, D. and Liu, S. 2018. Do analysts gain an informational advantage by visiting listed companies?. *Contemporary Accounting Research* 35 (4), pp. 1843-1867.

Hilary, G. and Hsu, C. 2013. Analyst forecast consistency. *The Journal of Finance* 68, pp. 271–297.

Hirshleifer, D. and Peng, L. and Wang, Q. 2021. News Diffusion in Social Networks and Stock Market Reactions. SSRN Working Paper.

Hirshleifer, D., Levi, Y., Lourie, B., Teoh, S. H. 2019. Decision Fatigue and Heuristic Analyst Forecasts. *Journal of Financial Economics* 133, pp. 83-98.

Hong, H., Kubik, J. D. and Solomon, A. 2000. Security analysts' career concerns and herding of earnings forecasts. *RAND Journal of Economics* 31, pp. 121–144.

Hope, O. K. 2003. Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research* 41 (2), pp. 235–272.

Hou, J., Zhao, S., and Yang, H. 2018. Security Analysts' Earnings Forecasting Performance Based on Information Transmission Network. *Physica A: Statistical Mechanics and Its Applications* 509, pp. 611-619.

Jacob, J., T. Z. Lys, and M. A. Neale. 1999. Expertise in forecasting performance of security analysts. *Journal of Accounting and Economics* 28 (1), pp. 51–82, https://doi.org/10.1016/S0165-4101(99)00016–6

Jegadeesh, N., J. Kim, S. Krische, and C. M. C. Lee. 2004. Analyzing the analysts: When do recommendations add value? *Journal of Finance* 59 (3), pp. 1083–1124.

Jiang, X. and Yuan, Q. 2018. Institutional investors' corporate site visits and corporate innovation. *Journal of Corporate Finance* 48, pp. 148-168.

Kar, N., & Kar, B. 2002. Social cognition and the participatory planning process. *Systems Research and Behavioural Science: The Official Journal of the International Federation for Systems Research* 19(4), pp. 377-382.

Kaustia, M. and Rantala, V. 2015. Social learning and corporate peer effects. *Journal of Financial Economics* 117, pp. 653–669.

Kumar, A. and Rantala, V. and Xu, R. 2022. Social Learning and Analyst Behavior. *Journal of Financial Economics* 143(1), pp. 434-461.

Lang, M. and Lundholm, R. 1996. Corporate Disclosure Policy and Analysts Behaviour. *The Accounting Review* 71, pp. 467-492.

Lang, M. H., and R. J. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (4), pp. 467–92.

Lim, T., 2001. Rationality and analysts' forecast bias. *The Journal of Finance* 56, pp. 369–385.

Malmendier, U. and Shanthikumar, D., 2014. Do security analysts speak in two tongues? *The Review of Financial Studies* 27, pp. 1287–1322.

Mikhail, M. B., B. R. Walther, and R. H. Willis. 1997. Do security analysts improve their performance with experience? *Journal of Accounting Research* 35 (Supplement), pp. 131–157. .

Moretti, E., 2011. Social learning and peer effects in consumption: evidence from movie sales. *Review of Economic Studies* 78(1), pp. 356-393.

Stickel, S. E. 1992. Reputation and performance among security analysts. *Journal of Finance* 47 (5), pp. 1811–1836.

Stock, J. H., Yogo, M. 2005. Testing for weak instruments in linear IV regression. Identification and inference for econometric models: Essays in honour of Thomas J. Rothenberg. *Cambridge University Press*, pp. 80–108.

Trueman, B. 1994. Analyst forecasts and herding behaviour. *The Review of Financial Studies* 7 (1), pp. 97-124.

Turkle, S. 2011. *Alone together: Why we expect more from technology and less from each other.* Basic Books.

Welch, I. 2000. Herding among security analysts. *Journal of Financial Economics* 58, pp. 369–396.

Zhao, Z., Zhang, Y., Feng, X. and Zhang, W. 2014. An analysis of herding behavior in security analysts' networks. *Physica A: Statistical Mechanics and its Applications* 413, pp. 116-124.

**Authorship contribution statement**

Mengjia Li: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing - original draft.

Wenjie Ding: Conceptualization, Validation, Writing- Reviewing and Editing, Funding acquisition, Supervision.

Hao Li: Conceptualization, Validation, Writing- Reviewing and Editing, Supervision.

Qingwei Wang: Conceptualization, Validation, Writing- Reviewing and Editing, Supervision.

Jason Zezhong Xiao: Conceptualization, Validation, Writing- Reviewing and Editing, Supervision.

**Acknowledgement**

Wenjie Ding would like to thank the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 72203244) for funding this research. Every author has equally contributed to this paper. Sun Yat-sen University is the first affiliation of this paper.