Politics and Specificity of Information: Evidence from Financial Analysts' Earnings Forecasts in a Relationship-based Economy

Forester Wong*

T.J. Wong*

Tianyu Zhang#

September 2018

First draft. Please do not cite or circulate without the authors' permission.

This paper examines how politics shapes the way specificity of financial analysts' information affects their forecast accuracy. We posit that in a relational economy, political information about firms' performance is highly specific because political influence is exerted on firms primarily through relationships. Using Latent Dirichlet Allocation (LDA), a topical modeling method, on a comprehensive sample of 87,332 reports of Chinese financial analysts from 2010 to 2015, we find that the proportion of the firm-specific (non-firm-specific) information in an analyst's report is associated with significantly higher (lower) earnings forecast accuracy relative to other analysts following the same firm in the same year. When political influence on firms increases, we find that firm-specific information is more positively associated with the analysts' relative forecast accuracy. We also find that the positive association between firm-specific information and relative forecast accuracy is stronger in provinces with more government intervention and weaker protection of property rights. Finally, we validate our LDA measures using the earnings component model in Ball and Brown (1967) and the stock return synchronicity model in Morck et al. (2000).

* Forester Wong and T.J. Wong are at University of Southern California
# Tianyu Zhang are at The Chinese University of Hong Kong

We thank the financial support of the Marshall iORB Outlier Research Grant and the technical support from the Corporate Governance Center at the Shenzhen Financial Institute. We also thank Wuyang Zhao for his comments and Yang Xuan for the provincial industrial policy data.

1. Introduction

In a relationship-based economy, firms not only contract with each other based on relational ties, politicians exert their influence on the firms through relationships as well (Rajan and Zingelas, 1998; Fisman, 2001; Allen et al., 2005; Faccio, 2006). The information about these relationships is highly firm-specific because each relationship is different. It is also highly value relevant because the performance and enforceability of the contracts depend on these relationships (Macneil, 1977; Williamson, 1979).[1]

The objective of the paper is to examine how politics shapes the way specificity of analysts' information affects their forecast accuracy. Since information of relationships is highly specific, we posit that in a relationship-based economy, analysts' firm-specific information is more associated with forecast accuracy than non-firm-specific information. Also, when the political influence on firms increases (e.g. due to changes in government contracts or policies), firm-specific information becomes even more important for forecasting because political relationships are highly value-relevant and firm-specific. This increase in forecasting power is likely to be due to political relationships being more important than other relationships (business ties) in a relational economy.

It is not clear whether financial analysts in a relational economy will specialize in gathering and disseminating firm-specific. On the one hand, financial analysts will develop their core competence through analyzing industry or market trends that are non-firm-specific since information about the relationships is not readily available in the market. Firms are less inclined to provide voluntary disclosures about relational contracts or political connections than arm's length contracts because relationships are often proprietary and hard to verify (Fan and Wong,

[1] Examples of how relationships are essential for evaluating firms' contracts and making forecasts are presented in the Appendix I.

2002; Ball et al., 2003; Li et al., 2018). The mandatory reporting of financial statements also does not contain information about these relationships due to the difficulty in measuring them. In addition, many of these relational economies are under strong government influence through its industrial and macro policies, analysts may specialize in gathering information about these policies that is non-firm-specific.

On the other hand, financial analysts will try to establish comparative advantage by gathering information about the firms' relationships that are highly firm-specific and essential for evaluating the firms' contracts. Since information pertains to the firms' relationships is imperative for generating earnings forecasts but it is in short supply, the analysts that possess such information will have the comparative advantage in forecasting. This also applies to information about government industrial and macro policies as well because firms that are closely connected with the politicians will receive more benefits.

To address these research questions, we use a comprehensive sample of 87,332 financial analysts' reports from China in 2010 to 2015. Only the latest analyst report issued before earnings announcement by each analyst each year is included in our final sample. We pick China for the study because it is the world's largest relationship-based economy with a vibrant equity and financial analyst markets. There are more than three thousand listed firms traded in the two domestic exchanges with a total of 3,175 unique analysts following these firms.

To measure the kind of information that gives Chinese financial analysts the competitive edge in forecasting, we examine the content of the analyst reports. We posit that the textual information contained in the analyst report should reflect the key information that is being used in the forecasts. The analysts have incentives to disseminate such information in the reports because it can signal that they have access to valuable information and provide their clients more information to interpret their forecasts. To operationalize on this, we have developed an objective method to identify textual content that contains relatively more firm-specific or non-firm-specific information. In particular, we use Latent Dirichlet Allocation (LDA hereafter), a topical modeling method, to identify 370 topics in the financial analysts' reports.[2] Using a set of objective criteria, we then classify these topics into four different categories: industry-specific topics, firm-specific topics, general topics and unclassified topics.

Our classification is done based on how these topics are distributed across firms. If a particular topic has been mentioned in 50% or more of the firms in a particular 1-digit SIC industry, we classify the topic as industry-specific. We consider a particular topic is mentioned in a firm if the topic has appeared in any of the firm's analyst report with at least 1% weight. This procedure yields 60 industry-specific topics. For the remaining 310 topics, we divide them into top and bottom terciles based on the number of firms whose analysts have mentioned the topics. The bottom tercile, which are topics mentioned in fewer firms, are considered firm-specific topics. Similarly, the top tercile of the topics are considered general topics. Topics that are either industry-specific or general are considered non-firm-specific topics.

We acknowledge that this approach only provides us with a relative measure of information specificity across topics. The relative distribution of these topics across the firms and industries in the sample can indicate the information specificity of the topics. That is, for topics that are distributed across a smaller number of firms, the content in such topics is likely to contain more firm-specific information. Similarly, for topics that are distributed in greater number of firms or concentrated within an industry, we expect that the content of these topics is likely to be at the industry or market levels, which is therefore less firm-specific. As will be discussed in Subsection

[2] Using the Coherence score technique, we identify that 370 LDA topics is the optimal number of topics for our sample of financial analysts' reports. See Subsection 3.2 for the discussion of the method.

4.6, we find qualitatively similar results in our main tests when we use different cutoffs for the firm-specific and general topic classifications.

We find results consistent with our conjecture that the firm-specific information provided in the reports gives financial analysts the competitive edge in forecasting earnings. The relative weight of the firm-specific topics is positively associated with the relative accuracy of the earnings forecasts in the reports. The relative accuracy is defined as the forecast accuracy in comparison with the average forecast accuracy by all other analysts for the same firm in the same year. Thus, this result suggests that compared with the contemporaneous cohort of analysts following the same firm, firm-specific information gives the analyst a comparative advantage in making the forecasts. However, we find non-firm-specific topics significantly *reduce* analysts' forecast accuracy.[3] This is inconsistent with the results of Piotroski and Roulstone (2004) that U.S. analysts develop core competence in identifying, gathering and disseminating common industry-level components. In China, however, since the emphasis on relationships in contracting may heavily influence firms' behavior in an industry or how politicians administer industrial policies, the specificity in these relationships may outweigh the common industry-level components in shaping the type of information analysts need for forecasting. Thus, Chinese analysts tend to specialize in producing firm-specific information in order to build their core competence.

Next, we test whether firms' experience of greater political influence will increase the positive association between firm-specific information and analysts' relative forecasts accuracy.[4] We use three approaches to address this question. First, we measure the political content of analyst

[3] One interpretation of this result is that by focusing too much on general or industry-specific information, analysts may misinterpret public information when making the forecasts. For example, without considering a firm's relationships, an analyst may extrapolate from past performance when in fact the firm's close relationship to an incoming politician may enable this firm to revert from prior losses to gains in future years.

[4] Past research has found that firms' connections with the government or politicians have significant impact on their performance and governance (Fan et al. 2007; Berkman et al., 2010; Chen et al., 2011; Piotroski et al., 2018).

reports using a word list of political words or the analyst reports' similarity in thematic content with the People's Daily, a Party newspaper that promotes government policies. We find that the positive association between firm-specific information and analysts' forecast accuracy is even stronger for analyst reports that contain more political content than less political content.

Second, we identify two different events that heightened political influence on a subset of firms in our sample. The first event relates to the turnover of provincial politician (governor or party secretary). When a province undergoes a politician turnover, firms that are headquartered in the province will experience a disruption in their political ties, which heightens the political effects on the firms. For these set of firms, we find that the relationship between firm-specific information and analysts' forecast accuracy is stronger in the two years after the politician turnover. The second event relates to the different industrial policies that are being promoted by local provinces. Provinces in China are highly decentralized in their industrial policies and support different sets of industries in their five-year plans.[5] When provincial leaders are promoting certain industries, the firms belonging to those industries will experience increased political influences. For these set of firms, we document that the relationship between firm-specific information and analysts' forecast accuracy is stronger during the years in which the underlying industry is being supported.

Third, we interact firm-specific information with two separate measures of the institutional environment of the province in which the firms operate. We find that firm-specific information is even more valuable to forecasting when the firms operate in provinces with greater government intervention or weak private property rights protection. These results provide supporting evidence

[5] Among the 19 industries (based on China's 2-digit SIC code) that are supported by at least one province, only 4 industries have policies that are supported by at least 30 of the 34 provinces (including provincial level administrative units). Ten of the 19 industries are supported by fewer than 10 provinces.

that politics in China increases the positive association between the specificity of financial analysts' information and forecast accuracy.

We provide two validity checks for the LDA classification. First, we correlate the financial analyst reports' topics with firms' earnings components based on the Ball and Brown (1967) model. We find that the average relative weight of the firm-specific topics of a firm is significantly positively associated with the level of its firm-specific component of the earnings innovation. However, the average relative weight of the non-firm-specific topics is not associated with the firm-specific component of the earnings innovation. Our second validation test is based on the stock return synchronicity measure in Morck et al. (2000). We find that firms with higher average relative weight of firm-specific topics in their analyst reports are associated with lower stock return synchronicity.

Finally, we use analyst departure as a shock to the firm's information environment, to demonstrate whether the information disseminated via the analysts has a causal effect on firm-specific information in the market. By showing that the firm-specific information can impact stock returns and increase the market's information environment, it provides credence to our main argument that the specific-information in the analyst reports has information value and can give analysts comparative advantage in forecasting. To perform the test, we compare the effect on a firm's stock return synchronicity by its departing analysts that produce higher level (top quartile) of firm-specific topics prior to its departure vs. those that produce fewer firm-specific topics (bottom quartile). We define the analyst's departure as the analyst no longer provides forecasts for any firms in the sample after the departure. Our results show that there is a more significant *increase* in a firm's 12-month stock return synchronicity after the departures of its analysts that produce more firm-specific topics than those that produce fewer firm-specific topics.

Our paper contributes to the literature in several ways. First, prior research finds that U.S. analysts value and benefit from private information from management (Brown et al., 2015), and especially through frequent interactions because of proximity (O'Brian and Tan, 2014) and social ties (Cohen et al., 2010). However, no study has specifically looked at whether it is the firm-specific or non-firm-specific information obtained privately from the firms that increase analysts' forecast quality. Second, past studies find that U.S. analysts' core competence is in their ability to provide industry-specific information to the market (Piotroski and Roulstone, 2004), and this ability gives them information advantage over firm managers (Hutton et al., 2012; Ali et al., 2017). We find that in a relationship-based economy like China, where information about firms' relationships is not readily available in public but essential for forecasting performance, analysts build their competitive edge by gathering and disseminating firm-specific information, not non-firm-specific information. More importantly, politics in China can increase the positive association between the specificity of financial analysts' information and forecast accuracy.

Second, this study adds to a growing literature on the role of information intermediaries in China. Li et al. (2018) find that in China private information on firms' relationships, which is otherwise hard to be disclosed, is disseminated to the public via analysts that share social ties with the firms. We argue that this private information is likely to be the firm-specific information and it provides them with a stronger comparative advantage in forecasting earnings than non-firm-specific information. Also, two recent papers, Piotroski et al. (2017) and You et al. (2017) use tone analysis and find that the corporate news in Chinese media is positively biased. In this paper, we use a topical modeling analysis to study the topics of financial analysts and find that the firm-specific information disseminated in their reports is giving them an edge in forecasting earnings and providing the market with firm-specific information.

Third, our results contribute to a growing accounting and finance literature that uses the LDA method in studying the content of firm disclosures and analyst reports. Huang et al. (2017) compare the similarity of the thematic content of the conference calls and subsequent analyst reports and examine if the analysts provide new information to the market. Using a similar approach, Bushman et al. (2017) study the content of banks' 10K business description and MD&A discussions, and find that a bank's systemic risk is positively associated with its similarity in topical content with other banks. In this study, instead of identifying how similar the documents (analyst reports and 10Ks) are based on their thematic content, we are interested in measuring how specific each topic is in the documents (analyst reports). That is, we consider a topic to be firm-specific (non-firm-specific) if it appears in the analyst reports of fewer (more) firms.

Fourth, this paper adds to the literature on stock return synchronicity pioneered by Morck et al. (2000). Our finding of the negative association between analysts' dissemination of firm-specific information and firms' stock return synchronicity provides support to their interpretation that low synchronicity is driven by firm-specific information but not noise. This interpretation is further supported by the result that the firm-specific information used in the study is associated with higher forecast accuracy.[6]

The remainder of the paper is organized as follows. Section II discusses the hypothesis development of the study. Our data and sample are discussed in Section III and results are presented in Section IV. We conclude our study in Section V.

2. Hypothesis Development

[6] Our results provide another explanation for why countries have low R^2. When protection of private property rights is weak, firms engage in relationship-based transactions. Since these relationships are proprietary and hard to verify, the firms provide very little public disclosures and reporting of these relationships to the market, which explains the firms' high stock return synchronicity. Moreover, our evidence shows that financial analysts play the role of gathering and disseminating this private firm-specific information to the market, thereby narrowing the gap in stock return synchronicity of relationship-based economies with market-based economies.

Firms operating in a relationship-based environment often turn to informal mechanisms for contracting since public laws and courts are ineffective against opportunism. These relational contracts increase firms' opacity because relationships are proprietary and hard to verify (Rajan and Zingeles, 1998; Li et al., 2018). Thus, firms do not make voluntary disclosures about these relationships to the market. This is consistent with the results in Morck et al. (2000) that markets with weak property rights protection have high R^2 or low stock return synchronicity (China is the second highest in the world in R^2), suggesting that firm-specific information is not readily available in these economies. The high R^2 indicates that the market prices have incorporated information that is more related to industry or market-level information, rather than firm-specific information.

In such an information environment, Chinese financial analysts cannot rely on publicly available firm-specific information for their analysis and formulating earnings forecasts. Even in the U.S. where there is more firm-specific information in the market (U.S. is ranked the highest in the world in R^2 in Morck et al., 2000), Piotroski and Roulstone (2004) find that financial analysts mainly provide industry or general information rather than firm-specific information to the market. This could be particularly true for China since the government embarks on a different set of industrial policies in each of the five-year plans. Like the U.S. financial analysts, their Chinese counterparts also build their core competence in analyzing industries by providing analyses on how a particular policy or a set of policies will impact a certain industry. When comparing with U.S. financial analysts, the Chinese analysts may even be more specialized in providing industry-specific information rather than firm-specific information to the market.

On the other hand, financial analysts may develop their comparative advantage by providing firm-specific information to the market in a relationship-based economy. Chinese analysts make transaction-specific investments such as paying frequent site visits to the firms

(Cheng et al., 2016) and/or building personal relationships with the firms (Li et al., 2018) and their other stakeholders. The analysts can accumulate a mosaic of information about the firms' relationships with their key stakeholders through their frequent interactions and communication with the firms. The transfer of the information of the firms' relationships can be done privately to protect its propriety. Over time, the analysts' relationships with the firms and their networks will enable them to verify the strength of their relationships. The transfer of this information does not violate China's Reg FD because it is not about a material event that can significantly move stock prices.[7] With this knowledge, however, the analysts can combine it with a seemingly immaterial piece of private or public information and make a more accurate forecast that they would not be able to do otherwise (see Appendix I for the four examples). Thus, we predict that analysts in relationship-based economies do provide information pertaining to the key relationships of the firms that they follow.[8] Since every relationship is unique, this kind of information is likely to be firm-specific.

With these arguments, we can reconcile our prediction with the results in Piotroski and Roulstone (2004). Chinese firms have few firm-specific information because it is hard to disclose the relationships associated with the relational contracts. However, Chinese analysts play the role as the embedded intermediaries by providing firm-specific information to the market. Thus, they help to reduce the synchronicity but not to the level of the U.S. because this way of firm-specific information dissemination is not as efficient as firms' direct voluntary public disclosures in the U.S. market. In the U.S., a high level of firm-specific information has been disclosed to the market by the firms. Analysts' comparative advantage is in identifying common information components

[7] China's Reg FD stated in China Securities Regulatory Commission Articles 40 and 128.
[8] The analysts will not reveal proprietary information about the firms' relationships but share information pertaining to the relationships. For example, they will talk about how certain government policy will benefit a firm without revealing details about the firm's close relationship with the government.

of the firm-specific information in the industry and how specific information of one firm affects another firm in the same industry. In China, since firm-specific information is less readily available and the industry's common information component is harder to identify because the effect of industrial policies depends on firms' political relationships, financial analysts build their core competence on collecting firm-specific information rather than industry-specific information.

Therefore, on the one hand, there are reasons to believe that the industry-specific and market-wide information gathered by the Chinese financial analysts give them the comparative advantage in forecasting because firms in general are very opaque in disclosing firm-specific information. On the other hand, the Chinese financial analysts' core advantage may be in disseminating firm-specific information to the market because it is difficult for the firms to directly disclose their relationships to the market. Our first hypothesis is as follows:

H1: The firm-specific information is more positively associated with the analyst's relative earnings forecast accuracy than the non-firm-specific information in an analyst's report.

Our second hypothesis is to examine if politics increases or decreases the association between the specificity of financial analysts' information and their forecast accuracy. We posit that when the political influence on firms increases, firm-specific information that is likely to capture such increase in political impact on firm performance will have a stronger association with forecast accuracy. We argue that this political information is likely to be firm-specific because the government exerts influence on firms via relationships. Even for firms that are subject to government's industrial policies in each five-year plan, the impact of the policies on each firm may depend on how the local politician is connected to the firm in her own locality. That is, a local government policy may support an industry but the real benefits may only go to certain firms that are connected to the local politician. Thus, a local industrial policy may not necessarily benefit all the firms in a particular industry simultaneously. Some firms may be hurt by the policy because favorable terms may only go to the connected firms and the unconnected firms' competitive strength may be weakened. To evaluate how an industrial policy will affect a firm in a particular region, the financial analysts will need to know the firms' relationship with the local politician that has the power to allocate resources.

To test this prediction, we use three different approaches to capture the amount of political influence on the firms at the time when the analyst report is produced. First, we measure the political content of the entire analyst report using a word list of political words.[9] We also use the similarity in thematic content between the analyst report and the political and economic sections of the People's Daily newspaper to gauge the amount of political content in the report. The People's Daily is a newspaper controlled by the Chinese central government which publishes content to promote government policies. The political content in the analyst reports captures the amount of political influence on the firms.

Our second approach directly identifies firm-years with high political influence. The first event relates to the turnover of provincial politician (governor or party secretary). When provinces undergo politician turnover, firms that are headquartered in those provinces will experience a disruption in the firms' political ties and increases the uncertainty of political effects on the firms. The second event relates to the different industrial policies that are being promoted by provincial leaders. Provinces in China are highly decentralized in their industrial policies and support different sets of industries in their five-year plans. When provincial leaders are promoting certain

[9] Our list of Chinese political words is based upon the *Dictionary of Scientific Development* (Xi 2007). This word list, which is used in Piotroski, Wong, and Zhang (2017) for measuring the political content of corporate news, contains Chinese political phrases and political slogans included in official Chinese Communist Party economic policy documents between 1978 and 2008, and was created by the government to celebrate 30 years of economic reforms.

industries, the firms belonging to those industries will experience increased government intervention.

Our final approach takes advantage of the variation in institutional differences across provinces in China. We posit that the firm-specific information is more positively associated with forecast accuracy in provinces with more government intervention or weak protection of property rights. This test will provide corroborating evidence that the effect of politics on analysts' information specificity is likely to be coming from government's political influence on firms. We use the first two approaches for the tests in Hypothesis 2a, and the third approach for Hypothesis 2b.

H2a: The positive association between firm-specific information and relative earnings forecast accuracy increases as the political influence on the firms increases.

H2b: The positive association between firm-specific information and relative earnings forecast accuracy is stronger in provinces with more government intervention and weaker property rights protection.

3. Sample and Empirical Measures

3.1 Sample

Our sample spans from January 2010 to December 2015. We obtain 196,243 unique analyst reports from Today Investment Co. Each of these reports usually contains multiple forecasts over different horizons or time periods. Therefore, we first remove all quarterly forecasts, and if there are still multiple annual forecasts remaining, we retain the forecast that is closest to the report date. This is based on the assumption that the content of each report is most relevant to the earnings of the year closest to the report's issuance date. For example, an analyst report filed on May 5, 2010, may contain forecasts for the year ending 2010, 2011, and 2012. Our selection procedure would limit our attention to the 2010 forecast only.

We begin with a sample of 2,703 unique firms. Since our main analyses focus on the relative forecast accuracy among analysts, we restrict our sample to the *latest* report issued before earnings announcement by each analyst for each firm in each year. It is important to note that this requirement has reduced the sample by more than half (from n=196,243 to 87,332).

We extract all forecast related data from Today Investment, which includes: Forecast EPS, Report Date, Brokerage Name, Brokerage Classification, Analyst Name, Company Name and Forecast Period. All other accounting data are obtained from The China Stock Market and Research Database (CSMAR). The analysts' external ranking (the star analyst status) is obtained from the *New Fortune* magazine. We manually match the analyst names in our sample with the *New Fortune* list—the match is made for the year prior to the year in which the *New Fortune* magazine was published.

3.2 Measuring analyst report content

We use Latent Dirichlet Allocation (LDA), one of the most popular topical-modeling techniques in textual analysis developed by Blei et al. (2003), to identify topics and their corresponding distributions within each analyst report. Past research has shown that LDA can meaningfully capture the topics of the textual content of analyst reports and 10-Ks (Bao and Datta, 2014; Dyer et al., 2017; Hoberg and Lewis 2017; Huang et al., 2017). The advantage of LDA topic modeling is that it is an unsupervised learning algorithm and hence does not require any labeled data to generate the topics. The LDA algorithm assumes that each document can be represented by a mixture of topics, and each topic also can be characterized by a probability distribution over the words. After specifying the number of topics, the algorithm will learn the probability distribution over all the words for each topic. More importantly, for each document, the algorithm will also identify the distribution of topics within it. We rely on the document-topic distribution discovered by LDA to identify the content of the analyst reports.

After extracting all the words within the 196,243 analyst reports, we do a number of steps to preprocess the documents. First, we remove all stop words, digits and lemmatization from our word corpus. Next, we remove all the words that either appear in more than 95% of our reports or in less than 3 articles. After this filtering procedure, we restrict our corpus to the 218,335 words that appear most frequently in our documents.

Using this corpus, we determine the optimal number of topics using Coherence score.[10] Starting from 100 topics, we calculated the Coherence score with 10 topics increments when the number of topics is less than 600 and 50 topics increments when the number of topics goes beyond 600 for concern of calculating and find that 370 topics provide the highest Coherence score. Therefore, we instruct the LDA algorithm to generate 370 topics and base our main results on 370 topics.[11] We have also set our topics to 200 and 500, and as discussed in Subsection 4.6, the results of the main test remain qualitatively the same.

3.3 Firm-Specific-Topic classification

As presented in Table 1 Panel A, our topical classification procedure yields a total of 103 firm-specific topics, 103 general topics and 60 industry-specific topics. Each topic is considered relevant to an analyst report or to a firm if it appears with a topical weight of at least 1% in the report. On average the firm-specific topics are relevant to ~78 firms, while the general topics are

---

[10] Researchers employ a variety of metrics such as perplexity or held-out likelihood to determine the optimal number of topics when using LDA. While, such measures are useful for evaluating the predictive model, they do not ensure the interpretability of the topic output. For example, Chang, Boyd-Graber, Gerrish, Wang and Blei (2009) find that these metrics (perplexity and held-out likelihood), which achieve better predictive perplexity, often cluster keywords from multiple unrelated ideas into one topic, resulting in outputs that are less interpretable. Since the objective of this paper is to identify topics containing firm-specific information, we want to prevent our topics from containing keywords that pretends to multiple unrelate ideas, hence we rely on Coherence score to identify the optimal number of topics.

[11] LDA topics also requires two hyper-parameters; alpha and beta. Alpha affects the sparsity of the document-topic distribution and beta affects the sparsity of topic-word distribution. Both default to a symmetric 1.0/number of topics.

relevant to ~314 firms and the industry specific topics are relevant to 123 firms (Panel B). While on average firm-specific topics account for ~45% of the analyst reports, and the general topics account for ~32% of the analyst reports and the industry topics accounts for ~11% of the analyst reports.

4. Empirical Results

4. Analyst comparative advantage

In this section, we first examine the type of information that creates the comparative advantage among financial analysts in making earnings forecasts. We then provide validation tests to show if our topical measures can capture firm-specific and non-firm-specific information. Finally, we use firm-specific-political topics to provide supporting evidence that information about firms' relationships is the source of analysts' comparative advantage in forecasting.

4.1 Measuring analyst comparative advantage

Forecast accuracy is one of the most important dimension along which analysts and their brokerage houses compete. Following Bae et al. (2008) we measure relative forecast accuracy as the proportional mean absolute forecast error. More specifically, we define relative forecast accuracy as the ratio of the difference between the absolute forecast error of analyst *i* of firm *j*'s in fiscal year *t* , $AFE_{i,j,t}$, and the average absolute forecast error across all analysts of firm *j*'s fiscal year *t*, $AvgAFE_{i,j,t}$, to the mean absolute forecast error. i.e.,

$$Relative\ Forecast\ Accuracy_{i,j,t} = \frac{AFE_{i,j,t} - AvgAFE_{j,t}}{AvgAFE_{j,t}} \tag{1}$$

A positive value for this variable indicates that the absolute forecast error of analyst *i* of firm *j* in fiscal year *t* is larger than the average absolute forecast error of all the forecasts of firm *j* in the same fiscal year.

4.2 What kind of information leads to analyst comparative advantage?

To examine which type of information leads to analyst comparative advantage, we estimate in Table 4 the following pooled analyst-level regression which controls for firm, brokerage and analyst characteristics:

$$\textit{Relative Forecast Error}_{i,j,t} = \alpha + \beta\ \textit{WFirmspecific}_{i,j,t} + \text{Control Variables} + \textit{FE}_{\textit{Firm},} \quad (2)$$

As mentioned above, we define *Relative Forecast Error*$_{i,j,t}$ as the proportional mean absolute forecast error of the *latest* earnings forecast by analyst i of firm j in year t. Our variable of interest, *WFirmspecific*$_{i,j,t}$, is the relative weight of firm-specific topics in the analyst report. A negative and significant coefficient on *WFirmspecific*$_{i,j,t}$ will indicate that firm specific information gives the financial analyst comparative advantage in forecasting earnings.

Since our measure of *Relative Forecast Error*$_{i,j,t}$ has already been de-meaned based on other forecast errors issued for the same firm during the same year, we include only the firm fixed effects and exclude the year fixed effects (however our main results remains similar when we include both firm and year fixed effects). To control for additional time-varying differences of the firm characteristics, we include several firm-specific control variables which may explain the content of the analyst reports. $BM_{j,t}$ is the book-to-market ratio of firm j measured at the beginning of the year t. $Size_{j,t}$ is the natural log of firm j's total market value at the beginning of the year t. $Loss_{j,t}$ is an indicator variable which takes the value of 1 if firm j is a loss firm in year t, and zero otherwise. Next, we include two market-based variables to control for the changes in the information environment. $Volume_{j,t}$ is the natural log of the annual trading volume in thousands of RMB (Chinese Yuan) for firm j and year t. $Stdret_{j,t}$ is the standard deviation of daily returns for firm j in the calendar year t. Finally, we include two additional variables to control for the demand that different clientele may have for analyst reports. *Institutions_Share*$_{j,t}$ is the percentage of institutional ownership and *%Foreign Analysts*$_{j,t}$ is the percentage of foreign analysts of firm j in year t. $Distance_{i,j,t}$ is an indicator variable which takes the value of 1 if firm j's headquarters is in the same province as analyst i's brokerage house in year t. $Foreign\_Ownership_{j,t}$ is the percentage of shares outstanding that is owned by foreign investors for firm j in year t.[12]

We also include several brokerage or forecast specific variables to control for differences in resources among the brokerage houses. $Horizon_{r,i,j,t}$ is the number of years between analyst i's forecast in report r for firm j and the firm's fiscal year end in year t. $Brokersize\_Analyst_{i,t}$ is the number of analysts hired by analyst i's brokerage for year t. $Brokersize\_Firm_{i,t}$ is the number of Chinese firms covered by analyst i's brokerage for year t.

On top of these, we use several analyst specific variables to control for analysts' ability. $Specialization_{i,t}$ is the number of different industries that analyst i covers in year t. $Experience\_Firm_{i,j,t}$ captures the firm-specific experience of analyst i and is measured as the number of years between the analyst's first forecast for firm j and the day of report r in year t. $Experience_{r,i,t}$ captures the overall experience of analyst i and is measured as the number of years between the analyst's first forecast for any firm in the database and the day of report r in year t. $Star_{i,t}$ is an indicator variable which takes the value of 1 if analyst i is nominated by the *New Fortune* Magazine as a star analyst in the year prior to year t. A summary of all the variable definitions is presented in Appendix II. The summary statistics of the control variables are reported in Table 2.

The results in column 1 of Table 3 show that the coefficient on $WFirmspecific_{i,j,t}$ is significantly negative, consistent with our conjecture that, analysts with more firm specific information have a higher comparative advantage. However, moving onto non-firm-specific

[12] Foreign ownership data are not available publicly, but Chinese firms disclose the identity and percentage of stock ownership of each firm's top 10 shareholders. We use this to proxy for the foreign ownership by summing the stock ownership of the top 10 owners that are foreign and treat the firms without top 10 owners that are foreign as firms with zero foreign ownership.

information, the coefficient on $WNonFirmspecific_{i,j,t}$, is positive and significantly different from zero, implying that having non-firm-specific information does not increase an analyst's comparative advantage. In addition, our results remain similar, when we decompose non-firm-specific topics into general ($WGeneral_{i,j,t}$) and industry topics ($WIndustry_{i,j,t}$) in Column 3 and including both firm-specific and non-firm-specific topical weights in the same regression (Columns 4 and 5).

4.3 Firm-Specific-Topic validation

We use two different methods that prior literature employs to capture firm-specific information to validate our topical classifications. First, similar to Piotroski and Roulstone (2004), we use stock return synchronicity to measure the amount of firm-specific, industry-level and market-level information impounded into stock prices. For each firm-year observation, we regress weekly returns on the current and prior week's equally-weighted market return ($MARET_{j,t}$) and the current and prior week's equally-weighted one-digit industry return ($INDRET_{j,t}$), or:

$$RET_{j,t} = \alpha + \beta_1 MARET_{j,t} + \beta_2 MARET_{j,t-1} + \beta_3 INDRET_{j,t} + \beta_4 INDRET_{j,t-1} + \varepsilon_{j,t} \quad (3)$$

We estimate this regression for each firm-year with 45 weekly observations and define synchronicity as:

$$SYNCH_{j,t} = \ln\left(\frac{R^2}{1-R^2}\right) \quad (4)$$

Lower value of $SYNCH_{j,t}$ indicates that a firm's stock returns are less closely tied with the market and industry and are assumed to reflect more firm-specific information. To test whether our classification reflects firm-specific and non-firm specific information, we estimate the following pooled firm-year model:

$$SYNCH_{j,t} = \alpha_0 + \beta_1 Avg\_Firmspecific_{j,t} + \beta_2 Fund\_Corr_{j,t} + \beta_3 Ln\left(Herf_{j,t}\right) + \beta_4 ST\_ROA_{j,t} + \beta_5 Ln\left(NIND_{j,t}\right) + \beta_6 Size_{j,t} + \sum_j^n Firm_{j,t} + \sum Year_t + \varepsilon_{j,t} \quad (5)$$

To control firm and year fixed effects, we include an array of firm and year indicator variables. To control for additional cross-sectional and time-varying differences, we include the correlation of the firm's earning with industry earnings ($Fund\_Corr_{j,t}$), the industry-level Herfindahl index ($Herf_{j,t}$) then volatility of the firm's earnings stream ($STD\_ROA_{j,t}$), the number of firms in the industry ($NIND_{j,t}$), and firm size ($Size_{j,t}$). After controlling for these attributes, we examine whether the average firm-specific topic weight ($Avg\_Firmspecific_{j,t}$), and the average non-firm-specific topic weight ($Avg\_NonFirmspecific_{j,t}$) across all reports issued for firm *j* in year *t*, is associated with stock return synchronicity. Our standard errors are clustered on firm and year.

As shown in Table 4, we find that when $Avg\_Firmspecific_{j,t}$ is high, firm *j*'s stock return synchronicity is significantly lower (column 1), implying that more firm-specific information is impounded into firm *j*'s stock prices. In contrast, when $Avg\_General_{j,t}$ is high (column 2), firm *j*'s stock return synchronicity is also significantly higher, implying that less firm-specific information is impounded into firm *j*'s stock prices. We observe similar results when we replace $Avg\_General_{j,t}$ with $Avg\_Industry_{j,t}$ (column 3).

Second, similar to Ball and Brown (1987) we use the amount of fundamental earnings that cannot be predicted by the market and industry earnings to identify the firm specific earnings component. More specifically, for each firm-year observation, we regress the change in ROA ($\Delta\text{ROA}_{j,t-\delta}$) on the change in quarterly market ROA ($\Delta\text{ROA}_{mk,t-\delta}$) and industry change in ROA ($\Delta\text{ROA}_{ind,t-\delta}$), or:

$$\Delta\text{ROA}_{j,t-\delta} = \alpha_{i,t} + \beta_{j,t}\Delta\text{ROA}_{mk,t-\delta} + \gamma_{j,t}\Delta\text{ROA}_{ind,t-\delta} + \varepsilon \quad (6)$$

We estimate this regression for each firm-year with 40 quarterly observations and use the estimated coefficients to calculate the expected change in ROA of firm *j* in period *t*, or:

$$E\left(\Delta\mathrm{ROA}_{j,t}\right) = \widehat{\alpha_{J,t}} + \widehat{\beta_{J,t}}\Delta\mathrm{ROA}_{mk,t} + \widehat{\gamma_{J,t}}\Delta\mathrm{ROA}_{ind,t} + \varepsilon \qquad (7)$$

We define the firm-specific earnings innovation as the absolute difference between the actual $\Delta\mathrm{ROA}_{j,t}$ and $E\left(\Delta\mathrm{ROA}_{j,t}\right)$, or:

$$FirmSpecEarnings_{j,t} = abs\left|\Delta\mathrm{ROA}_{j,t} - E\left(\Delta\mathrm{ROA}_{j,t}\right)\right| \qquad (8)$$

Higher value of $FirmSpecEarnings_{j,t}$ indicates more firm-specific earnings innovation. To test whether our classification reflects firm-specific and non-firm-specific information, we replace $SYNCH_{j,t}$ in equation (5) with $FirmSpecEarnings_{j,t}$:

$$FirmSpecEarnings_{j,t} = \alpha_0 + \beta_1 Avg\_Firmspecific_{j,t} + \beta_2 Fund\_Corr_{j,t} + \beta_3 Ln\left(Herf_{j,t}\right) + \beta_4 ST\_ROA_{j,t} + \beta_5 Ln\left(NIND_{j,t}\right) + \beta_6 Size_{j,t} + \sum_j^n Firm_{j,t} + \sum Year_t + \varepsilon_{j,t} \qquad (9)$$

As demonstrated in column 4, the coefficient on $Avg\_Firmspecific_{j,t}$ is positive and significant, implying that the average firm-specific topical weight is positively associated with the amount of firm-specific earnings innovation. In contrast, the coefficient on $Avg\_General_{j,t}$ and $Avg\_Industry_{j,t}$ is insignificantly different from zero (columns 5 and 6).

4.4 Does political influence alter the relationship between firm-specific information and analyst forecast accuracy?

4.4.1 Using political content within analyst reports to capture political influences

In this subsection, we use the amount of political content within a firm's analyst reports to proxy for the amount of political influence. We use two different ways to measure political content in an analyst report. First, we measure political content using $Political\ Words_{i,j,t}$, the percentage of political words within the documents.[13] Second, we use the political and economic sections of the People's Daily newspaper as a benchmark for political content. The People's Daily is a newspaper controlled by the Chinese central government which publishes content to promote government policies. We use two methods to quantify the amount of overlap in content between the People's Daily newspapers and the analyst's reports.

Our first method is based on the Tf-idf cosine similarities (term frequency-inverse document frequency). Each document is represented as a vector of size V, where V is the size of vocabulary, i.e. the number of unique words appear in the corpus. Similarities of two documents (vectors) can be measured by:

$$Similarity_{i,j,d} = \cos(\theta) = \frac{\boldsymbol{A}.\boldsymbol{B}}{\|\boldsymbol{A}\|\|\boldsymbol{B}\|} \tag{10}$$

, where **A** stands for the vectors of words in analyst *i*'s report for firm *j* in year *t* and **B** stands for the vectors of words in an article in the people's daily newspaper in day d.

$Cosin\_ Similarity_{i,j,t}$ is the median value of all the $Similarity_{i,j,d}$ value between analyst *i*'s report for firm *j* in year *t* and all the articles in the political and economic sections of The People's Daily between year *t-1* to year *t+1*.

Our second method is based on the distance between the LDA topics in the analyst's report and the newspaper. Using LDA we calculate each document's topic distribution and compare the topic distribution between the analyst's report and the newspaper. We calculate the similarity between LDA topic as:

$$LDASimilarity_{i,j,d} = -\sum_{l=1}^{N} \frac{(x_l - y_l)^2}{(x_l + y_l)} \tag{11}$$

[13] Our list of Chinese political words is based upon the *Dictionary of Scientific Development* (Xi 2007). This word list, which is used in Piotroski, Wong, and Zhang (2017) for measuring the political content of corporate news, contains Chinese political phrases and political slogans included in official Chinese Communist Party economic policy documents between 1978 and 2008, and was created by the government to celebrate 30 years of economic reforms.

, where x_l is the topic distribution of topic l for analyst i's report for firm j in year t and y_l is the topic distribution of topic l for an article in The People's Daily newspaper article on day d.

$LDA\_Similarity_{i,j,t}$ is the median value of all the $LDASimilarity_{i,j,d}$ value between analyst *i*'s report for firm *j* in year *t* and all the articles in the political and economic section of The People's Daily between year *t-1* to year *t+1*.

In Table 5, $HighPolitical1_{j,t}$ is an indicator variable which takes on the value of 1 when the average $Cosin\_ Similarity_{i,j,t}$ across all of firm j's analyst reports is higher than the median across all firms. Similarly, $HighPolitical2_{j,t}$ and $HighPolitical3_{j,t}$ are indicator variables which takes on the value of 1 when $LDASimilarity_{i,j,t}$ and $Political\ Words_{i,j,t}$ across all of firm j's analyst reports are higher than the median across all firms, respectively. We then interact these three indicator variables with $WFirmspecific_{i,j,t}$ to identify topics that are likely to both political and firm-specific. In Columns (1) to (3), we find that the relative forecast error is significantly lower when both firm specific topics and political content are higher. This support for hypothesis that political influences increases the importance of firm-specific information.

4.4.2 Politician turnover and provincial level industry policies

Using the two events discussed in Section 2, we identified a subset of firms-year in which the underlying firms are experiencing increased political influences. In Table 6, $Turnover_{j,t}$ is an indicator variable which takes on the value of 1 if firm *j* is headquartered in the province in which there is a key politician (secretary or mayor) turnover in year *t* or *t-1*. $Policy_{j,t}$ is an indicator variable which takes on the value of 1 if firm *j* belongs to an industry (2-digit SIC) that is being promoted by firm *j*'s headquarter province in year *t*.

In Column (1) when we interact $WFirmspecific_{i,j,t}$ with $Turnover_{j,t}$ and in Column (2) when we interact $WFirmspecific_{i,j,t}$ with $Policy_{j,t}$, we find both interaction terms are negative and significant, indicating that for these firms, firm-specific information is more important during these periods with heightened political influence. Overall, our results suggest that the relationship between firm-specific content and forecast accuracy can be significantly moderated by the level of political influence.

4.4.3 State level government interventions and property rights protection

In this subsection, we provide further evidence that the effect of analysts' information specificity has on forecast accuracy is likely to be coming from government's political influence on firms. The level of government interventions and property rights protections varies significantly between provinces. We use two indicator variables to capture this cross province institutional variation. $HighGovInt_{j,t}$ is an indicator variable which takes on the value of 1 if $GovIntervention_{j,t}$ is above median across all provinces in year *t*. $LowPptRight_{j,t}$ is an indicator variable which takes on the value of 1 if $PptRight_{j,t}$ is below median across all provinces in year *t*.For firms that are within provinces with either high level of government intervention or poor property rights protection, we expect the firm-specific content to play an even more significant role in the analyst's comparative advantage in forecasting earnings.

Table 6 Columns 3 and 4 shows that firm-specific information is even more significant in provinces where either government intervention is high, or property rights protection is weak. This result supports the conjecture that in provinces where government plays a bigger role, firm-specific information is providing an even stronger comparative advantage to the financial analysts to forecast.

4.5 Additional Tests

4.5.1 Do Chinese analysts provide firm-specific information to the market?

Our earlier test in Table 4 is consistent with our conjecture that analysts are providing firm-specific information to the market. However, the content inside an analyst's report is not randomly determined. Our earlier findings may be driven by the fact that analysts are simply including more firm-specific content when there is more firm-specific information in the market as reflected by the lower stock price synchronicity. We address this concern by investigating firms that experience one or more analyst/s departure. We define an analyst as having departed, if the analyst stops providing any forecast in our dataset. To ensure that we are capturing meaningful departures, we also require the analyst to have existed in our dataset for at least two years before departure. As shown in Table 7 Panel A, during our sample period, there is a total 1,068 analyst departures, and these departures are evenly distributed across 2011, 2012 and 2013.[14] We do not include any departures in year 2014 and 2015 because we cannot determine if the analysts have actually departed or are merely temporarily missing from our dataset.

Our departure event sample includes all firms that have at least one departing analyst. Among all the departing analysts, we classify them into either firm-specific or non-firm-specific analysts. To do that, we first rank all the analysts based on the amount of firm-specific information in all their reports. We then classify an analyst to be a firm-specific (non-firm-specific) analyst if he/she is ranked in the top (bottom) quartile. We exclude from our regression departures of analysts that do not belong to these two extreme quartiles. This regression compares the effect that a firm-specific analyst's departure has on post-departure stock return synchronicity, relative to that of the departure of a non-firm-specific analyst. The test is akin to a difference-in-differences model where we use the non-firm-specific analysts' departure to control for unobserved factors that may

[14] We stop our departure year in 2013 to ensure that each departing analyst has left the profession for at least two years.

confound the departing sample. The departure of firm-specific analysts is the treatment group. Equation (12) shows the regression model for the departure test:

$$SYNCH_{j,t} = \alpha_0 + \beta_1 \text{Departure}_{i,t} \times FirmSpecAnalyst_i + \beta_2 FirmSpecAnalyst_i + \beta_3 \text{Departure}_{i,t} + \beta_4 Fund\_Corr_{j,t} + \beta_5 Ln\left(Herf_{j,t}\right) + \beta_6 ST\_ROA_{j,t} + \beta_7 Ln\left(NIND_{j,t}\right) + \beta_8 Size_{j,t} + \sum_j^n Firm_{j,t} + \sum Year_t + \varepsilon_{j,t} \quad (12)$$

The unit of analysis is firms with a departing analyst (*i*)- pre-/post-(*t*). A firm is considered to have a departing analyst if the analyst issued the last earnings forecast for the firm during the year that the analyst departs. We limit the pre- (post-) departure observations included in the regression to within one year of the analyst's departure. We calculate the pre-departure $SYNCH_{j,t}$ starting 395 days and ending 30 days before the analyst departure. We calculate the post-departure $SYNCH_{j,t}$ starting 30 days and ending 395 days after the analyst's departure.

$FirmSpecAnalyst_i$ is an indicator variable that takes a value of 1 if departure event involves a firm-specific-analyst, and zero otherwise. $Departure_{i,t}$ is an indicator variable that takes a value of 1 for observations after the departure event, and zero otherwise. Thus, the $Departure_{i,t}$ variable captures the mean changes in a firm's synchronicity following the departure of an analyst, i.e., 365 days following the departure. The interaction term $Departure_{i,t} \times FirmSpecAnalyst_i$ is our main variable of interest; it captures the incremental effect of an analyst's departure when the departure involves a firm-specific-analyst. We predict that the increase in synchronicity following a firm-specific-analyst's departure will be greater than that of the non-firm-specific analyst. As before, we include in the estimation the year fixed effect (based on the departure event year) and the firm fixed effect. We include all control variables, defined in the Appendix II.

As demonstrated in Table 7, our results show that there is a significant *increase* in a firm's 12-month stock return synchronicity after the departure of its analyst that produces more firm-specific topics than those that produce fewer firm-specific topics. This suggests that our earlier finding that Chinese analysts are providing firm-specific information to the market is likely to be causal. In Column 4, we replace the dependent variable ($SYNCH_{j,t}$) in equation 12 with $FirmSpecificEarnings_{j,t}$. We find that the coefficients on $Departure_{i,t} \times FirmSpecAnalyst_i$ and $Departure_{i,t}$ are both insignificantly different from zero.[15] This suggests that the analyst's departure is unlikely to be related to any change in the specific component of earnings innovation of the firms.

4.5.2 Robustness tests

In this subsection, we discuss the results of a few robustness tests for the main tests. First, we use various partitions to separate the topics into firm-specific and general topics: top and bottom 50, 77 (quartiles) and 120 topics. We repeat the regressions in Table 3 using these new classifications for firm-specific and general information. Our results show that the coefficient on $WFirmspecific_{i,j,t}$ is significantly negative at the one percent level using the top 77 and 120 topics as partition, but the significance level drops to 10% level when using the top 50 topics as partition. None of the coefficient on $WGeneral_{i,j,t}$ is significantly different from zero using the three partitions for general topics.

Second, instead using the optimal number of 370 topics determined by the Coherence score, we use 200 and 500 topics as diagnostic checks. We repeat the regressions in Table 3 using both numbers of topical classifications and continue to find that the coefficient on $WFirmspecific_{i,j,t}$ to

[15] We include the entire sample of departing analysts and use above and below median of firm-specific information to separate the analysts into firm-specific (non-firm-specific) analysts. The coefficient on $Departure_{j,t} \times FirmSpecAnalyst_j$ remains positive though the significance weakens to the 10% level.

be negative and significant at the one percent level. However, the coefficients on $WGeneral_{i,j,t}$ and $WIndustry_{i,j,t}$ are not different from zero when 200 or 500 is used as the number of classification.

5. Conclusion

This paper finds that in a relationship-based economy such as China, it is the firm-specific information in the financial analysts' reports that give them the comparative advantage in forecasting earnings than non-firm-specific information. More importantly, we find that the level of political influence on a firm increases the positive association between firm-specific information and forecast accuracy. That is, when politics has a stronger influence on a firm, analysts will need more firm-specific such as information about the relationship between the politician and firm, to make their earnings forecasts. Our topical measures of firm-specific and non-firm-specific information are validated by the Ball and Brown (1967) earnings component model and Morck et al. (2000) stock return synchronicity model.

Future research should study whether and how relationship contracting affects analysts' knowledge acquisition and development of expertise or specialization. Analysts in these economies are likely not to build their competence on finding industry-specific components that drive performance, but rather focusing on gathering private information about firms' relationships and their impact on contracts. Their knowledge to acquire information and analyze the firms is not likely to be transferrable within industry boundary but among firms that share similar network ties. We should examine whether and under what conditions relational contracts determine how financial analysts or other information intermediaries choose their specialization.

References

Allen, F., Qian, J. and Qian, M., 2005. Law, finance, and economic growth in China. *Journal of financial economics*, *77*(1), pp.57-116.

Ali, A., Amiram D., Kalay A., and Sadka G., 2017. Industry expertise and the informational advantages of analysts over managers. Working paper. UT Dallas.

An, H., Y. Chen, D. Luo, and T. Zhang. 2016. Political uncertainty and corporate investments: Evidence from China. *Journal of Corporate Finance* 36: 174-189.

Brown, P. and Ball, R., 1967. Some preliminary findings on the association between the earnings of a firm, its industry, and the economy. *Journal of Accounting Research*, pp.55-77.

Ball, R., Robin, A. and Wu, J.S., 2003. Incentives versus standards: properties of accounting income in four East Asian countries. *Journal of accounting and economics*, *36*(1-3), pp.235-270.

Berkman, H., Cole, R.A. and Fu, L.J., 2010. Political connections and minority-shareholder protection: Evidence from securities-market regulation in China. *Journal of Financial and Quantitative Analysis*, *45*(6), pp.1391-1417.

Bushman, R., Chen, J. and Williams, C., 2017. Bank Connectedness: Qualitative and Quantitative Disclosure Similarity and Future Tail Co-movement. UNC, Working Paper.

Chen, C.J., Li, Z., Su, X. and Sun, Z., 2011. Rent-seeking incentives, corporate political connections, and the control structure of private firms: Chinese evidence. *Journal of Corporate Finance*, *17*(2), pp.229-243.

Chen, Z., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M. and Ghosh, R., 2013, October. Discovering coherent topics using general knowledge. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 209-218). ACM.

Cheng, Q., Du, F., Wang, X. and Wang, Y., 2016. Seeing is believing: analysts' corporate site visits. *Review of Accounting Studies*, *21*(4), pp.1245-1286.

Christensen, D. M., M. B. Mikhail, B. R. Walther, and L. A. Wellman. 2016. From K street to wall street: Political connections and stock recommendations. *The Accounting Review*, 92(3): 87-112.

Faccio, M., 2006. Politically connected firms. *American economic review*, *96*(1), pp.369-386.

Fan, J.P. and Wong, T.J., 2002. Corporate ownership structure and the informativeness of accounting earnings in East Asia. *Journal of accounting and economics*, *33*(3), pp.401-425.

Fan, J. P., T. J. Wong, and T. Zhang. 2007. Politically connected CEOs, corporate governance, and Post-IPO performance of China's newly partially privatized firms. *Journal of Financial Economics*, 84(2): 330-357.

Huang, A.H., Lehavy, R., Zang, A.Y. and Zheng, R., 2017. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*.

Hutton, A.P., Lee, L.F. and Shu, S.Z., 2012. Do managers always know better? The relative accuracy of management and analyst forecasts. *Journal of Accounting Research*, *50*(5), pp.1217-1244.

Julio, B., and Y. Yook. 2012. Political uncertainty and corporate investment cycles. *The Journal of Finance* 67(1): 45-83

Li, Z., Wong, T.J., Yu, G. 2017. Information dissemination through embedded financial analysts: evidence from China. Working paper. USC.

Macneil, I.R., 1977. Contracts: Adjustment of long-term economic relations under classical, neoclassical, and relational contract law. *Nw. UL Rev.*, *72*, p.854.

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

O'Callaghan, D., Greene, D., Carthy, J. and Cunningham, P., 2015. An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, *42*(13), pp.5645-5657.

Piotroski, J.D. and Roulstone, D.T., 2004. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. *The Accounting Review*, *79*(4), pp.1119-1151.

Piotroski, J.D., Wong, T.J. and Zhang, T., 2015. Political incentives to suppress negative information: evidence from Chinese listed firms. *Journal of Accounting Research*, *53*(2), pp.405-459.

Piotroski, J.D., Wong, T.J. and Zhang, T., 2017. Political Bias in Corporate News: The Role of Conglomeration Reform in China. *The Journal of Law and Economics*, *60*(1), pp.173-207.

Piotroski, J.D., Wong, T.J., and Zhang, T. 2018. Political networks and stock price co-movement: evidence from network-connected firms in China. Working paper. USC.

Williamson, O.E., 1979. Transaction-cost economics: the governance of contractual relations. *The journal of Law and Economics*, *22*(2), pp.233-261.

You, J., Zhang, B. and Zhang, L., 2017. Who captures the power of the pen? *The Review of Financial Studies*, *31*(1), pp.43-96.

Zhang, H., Li, L., Zhou, D. and Zhou, P., 2014. Political connections, government subsidies and firm financial performance: Evidence from renewable energy manufacturing in China. *Renewable Energy*, *63*, pp.330-336.

Appendix I

Below are the four cases that demonstrate how information of a firm's relationship can affect an analyst's ability to make earnings forecasts. The cases are based on interviews with Chinese analysts conducted by the authors of Li, Wong and Yu (2018). All four scenarios involve firms not being able to reveal proprietary information about their relationships with a supplier (case 1), among two large shareholders (case 2) and with the government (cases 3 and 4). However, the analysts have accumulated mosaic information about the firms' relationships over numerous interactions and conversations with the firms. The transfer of such information is not considered to be significant inside information because it does not have a direct effect on share value. But with such knowledge, connected analysts are better able to combine it with other new information to formulate more accurate earnings forecasts.

Case 1:

The firm is in tourism business. One of its core competences is its ability to secure from airlines outbound charter flights for popular routes that they can generate high volume of sales. Due to keen competition, the firm has to rely on its social ties to secure charter flights for popular routes from these airlines. For example, the firm has arranged charter flights with a Russian airline to serve customers that want to attend the 2018 World Cup. Without close social ties with this firm, an analyst is unable to know the firm's relationship with the airlines and its ability to secure the charter flights.

Case 2:

A company purchased another company through a stock swap. The owner of the target firm became a senior executive and large (not controlling) shareholder of the merged firm. It was well known to the market that the owner of the target firm had business opportunities that could enhance the merged firm's profitability. However, the connected analyst, knew that there was tension in the relationship between the owners of the target and the acquiring firms through her relationship with the target firm's owner. The opportunities of the target firm would more likely go to the merged firm when the owner of the target firm was given more ownership control of the merged firm. Thus, the connected analyst based her earnings forecast on the level of change in ownership control of the merged firm by the target firm's owner.

Case 3:

This central SOE is in the tourism industry. According to government policies, all the renewal of operating contracts for airport duty free shops needs to go through a bidding process. All qualified companies are invited to submit an open bid for the contracts. However, during the time when the Shanghai Airport was holding the auction for its duty free shops in 2018, the central SOE's connected analyst was correct in predicting that it would win the bid and the bidding price would be lower than its winning bid in the Beijing Airport auction held in 2017. The analyst's prediction was based on her private knowledge of the government's strong preference in the Shanghai auction. First, through her close relationship with the firm, she received information that the government had a strong preference to award the duty free shop contracts of major airports to this central SOE.

Second, she also found out that a certain central government agency had expressed this preference to other qualified bidding companies which she inferred would reduce the competition and lower the winning bid. Her thought was that firms would tend to listen to the government even though it was simply expressed as a preference and not an order. With this private knowledge, the connected analyst could make a more accurate prediction for the outcome of the Shanghai auction.

Case 4:

A large listed brokerage house was caught violating trading rules and was subject to harsh penalties in 2013. Its share price and profitability dropped to the lowest level in its sector. However, a connected analyst provided a forecast that was different from other analysts because she understood that the brokerage firm’s controlling shareholder had maintained its strong ties with the government and there was no threat of significant turnover in its senior management. Her strong social ties with the brokerage house enables her to draw conclusion that the trading rule violation had not seriously affected the firm’s relationship with the government. One year later, the brokerage firm emerged to be the leader in the sector in stock price performance, confirming the forecast of the connected analyst.

Appendix II: Variable Definition	
Wfirmspecific$_{i,j,t}$	Weight of firm-specific topics in percent for the report of analyst *i* of firm *j* in year *t*.
Wnonfirmspecific$_{i,j,t}$	Weight of non-firm-specific topics in percent for the report of analyst *i* of firm *j* in year *t*.
Wgeneral$_{i,j,t}$	Weight of general topics in percent for the report of analyst *i* of firm *j* in year *t*.
Windustry$_{i,j,t}$	Weight of industry-specific topics in percent for the report of analyst *i* of firm *j* in year *t*.
Political Words$_{r,i,j,t}$	Percentage of political words for the report of analyst *i* of firm *j* in year *t*.
BM$_{j,t}$	Book-to-market ratio for firm *j*, measured at the beginning of year *t*.
Size$_{j,t}$	Log (firm *j*'s total market value at the beginning of year *t*).
Volume$_{j,t}$	Log (annual trading volume in thousands of RMB) for firm *j* in year *t*.
Stdret$_{j,t}$	Standard deviation of daily returns for firm *j* for the calendar year *t*.
Horizon$_{r,i,j,t}$	The number of year(s) between analyst *i*'s forecast from the report of analyst *i* of firm *j*'s fiscal year-end in year *t*.
Experience_Firm$_{i,j,t}$	Number of year(s) between the analyst *i*'s first forecast for firm *j* and the day of the current forecast in year *t*.
Experience$_{r,i,t}$	Number of year(s) between the analyst *i*'s first forecast of any firm in the database and the day of the current of report r in year *t*.
Brokersize_Analyst$_{i,t}$	Total number of analysts hired by analyst *i*'s brokerage for year *t*.
Brokersize_Firms$_{i,t}$	Total number of firms covered by analyst *i*'s brokerage for year *t*.
Institutions_share$_{j,t}$	Ownership percentage (in the last quarter) of institutional investors (e.g., mutual funds, foreign institutional investors,

	brokerage firms, insurance companies, pension funds, investment trusts, and banks) of firm *j* in year *t*.
Specialization$_{i,t}$	The number of different industries that the analyst *i* covers during year *t*.
Star$_{i,t}$	Indicator variable takes on the value of 1 if analyst *i* is nominated by the *New Fortune* magazine as a star analyst in the year prior to year *t*.
% Foreign Analysts$_{j,t}$	The percentage of foreign analysts following firm *j* in year *t*
Loss$_{j,t}$	Indicator variable takes on the value of 1 if the firm *j* reports a loss in year *t*.
Foreign_Ownership$_{j,t}$	Indicator variable takes on the value of 1 if the underlying firm *j* shares the same province as analyst *i*'s brokerage house in year *t*
Foreign_Ownership$_{j,t}$	Percentage of shares outstanding that is owned by foreign investors for firm *j* at the beginning of the calendar year *t*
Relative Forecast Error$_{i,j,t}$	Following Bae et al. (2008) we measure forecast accuracy using the proportional mean absolute forecast error. More specifically, it is the ratio of the difference between the absolute forecast error of analyst *i* forecasting for firm *j*'s fiscal year *t* earnings and the average absolute forecast error across all analyst forecasts of firm *j*'s fiscal year *t* earnings, to the mean absolute forecast error. i.e., Relative Forecast Accuracy= $\frac{AFE_{i,j,t}-AvgAFE_{j,t}}{AvgAFE_{j,t}}$ A positive value for this variable indicates that the absolute forecast error of analyst *i* for firm *j*'s fiscal year *t* is larger than the average absolute forecast error of all the forecasts for firm *j* for the same fiscal year.
Synch$_{j,t}$	We estimate firm-specific measures of return synchronicity using the methodology outlined in Piotroski and Roulstone (2004). Specifically, for each firm-year observation, we regress weekly returns on the current and prior week's equally-weighted market return (*MARET$_{j,t}$*) and the current and prior week's equally-weighted one-digit industry return (*INDRET$_{j,t}$*), or:

	$$RET_{j,t} = \alpha + \beta_1 MARET_{j,t} + \beta_2 MARET_{j,t-1} + \beta_3 INDRET_{j,t} + \beta_4 INDRET_{j,t-1} + \varepsilon_{j,t}$$ Following Piotroski and Roulstone (2004), we estimate this regression for each firm-year with a minimum of 45 weekly observations and define synchronicity as: $$SYNCH_{j,t} = \ln(\frac{R^2}{1-R^2})$$ Lower value of $SYNCH_{j,t}$ indicates that a firm's stock returns are less closely tied with the market and industry and are assumed to reflect more firm specific information.
Fund_Corr$_{j,t}$	The correlation of a firm's earnings with industry level earnings. More specifically, we estimate the following regression: $$ROA_{j,t} = \alpha + \beta_1 INDROA_{j,t} + \beta_2 INDROA_{j,t-1} + \varepsilon_{j,t}$$ We estimate this regression for each firm-year using a minimum of 12 quarterly observations and define the fundamental correlation as the R^2 from this regression.
Herf$_{j,t}$	Industry concentration measured as the one-digit industry's Herfindahl index for industry of firm *j* in year *t*.
STD_ROA$_{j,t}$	Standard deviation of the quarterly return on assets for firm *j*, measured over three years preceding year *t*.
NIND$_{j,t}$	The number of firms in the industry (based on 1-digit industry code) of firm *j*.
FirmSpecEarnings$_{j,t}$	Similar to Ball and Brown (1967), we estimate the firm-specific component of earnings innovation using a three-step process: 1. For each firm-year observation, we estimate the following regression using 40 quarterly observations $$\Delta\text{ROA}_{j,t-\delta} = \alpha_{i,t} + \beta_{j,t}\Delta\text{ROA}_{mk,t-\delta} + \gamma_{j,t}\Delta\text{ROA}_{ind,t-\delta} + \varepsilon$$

	2. Then we estimate the expected change in firm j's change in ROA in period t based on the market and industry change in ROA. $E(\Delta\text{ROA}_{j,t}) = \widehat{\alpha_{J,t}} + \widehat{\beta_{J,t}}\Delta\text{ROA}_{mk,t} + \widehat{\gamma_{J,t}}\Delta\text{ROA}_{ind,t} + \varepsilon$ 3. The firm specific earnings innovation is the absolute different between the actual $\Delta\text{ROA}_{j,t}$ and $E(\Delta\text{ROA}_{j,t})$, or: $FirmSpecEarnings_{j,t} = abs\left\|\Delta\text{ROA}_{j,t} - E(\Delta\text{ROA}_{j,t})\right\|$; i.e. the higher this number, the more firm-specific is the earnings innovation.
$\text{Departure}_{i,j,t}$	Indicator variable which takes on the value of 1 if anyone of firm j's analysts has departed in the previous year.
$FirmSpecAnalyst_i$	Indicator variable takes on the value of 1 if the average $Wfirmspecific_{i,j,t}$ across all of analyst i's reports is ranked in the top quartile among all departing analysts.
$\#\ \text{Analyst}_{j,t}$	The number of analysts that issued a report for firm j in year t.
$HighPolitical1_{j,t}$	An indicator variable which takes the value of 1 if the average $Cosin\_Similarity_{all,\ i,j,t}$ across all of firm j's analyst reports is higher than median and zero otherwise. Where $Cosin\_Similarity_{all,\ i,j,t}$ is the Tf-idf Cosine similarities between analyst i's report issued for firm j in year t and the political section of People's Daily newspaper between year $t-1$ and year $t+1$.
$HighPolitical2_{j,t}$	An indicator variable which takes the value of 1 if the average $LDA\_Similarity_{all,\ i,j,t}$ across all of firm j's analyst reports is higher than median and zero otherwise. Where $LDA\_Similarity_{all,\ i,j,t}$ is the LDA topic distance between analyst i's report issued for firm j in year t and the political section of the People's Daily newspaper between year $t-1$ and year $t+1$.
$HighPolitical3_{j,t}$	An indicator variable which takes the value of 1 if

	Political_Words$_{i,j,t}$ across all of firm j's analyst reports is higher than median and zero otherwise. Where *Political_Words*$_{i,j,t}$ is the percentage of political words within report *r* of analyst *i* of firm *j* in year *t*.
Turnover$_{j,t}$	An indicator variable which takes the value of 1 if firm j's headquarter province experienced a key politician turnover during year t-1 or year t.
Policy$_{j,t}$	An indicator variable which takes the value of 1 if the industry in which firm j belongs to (same two digits SIC) is being promoted by firm j's headquarter province in year t.
GovIntervention$_{j,t}$	Index for the level of state government intervention in the province in the year. Higher value indicates less government intervention within that particular province. It is a sub-index of marketization index compiled by Fan, Wang and Yu (2016). Specifically, this index includes three components relating to the relationship between government and market, role of market in resource allocation, reduction of government intervention, and reduction of government size. Our data spans from 2010 to 2014, for year 2015, we fill in the value for 2015 using data from the previous year.
PptRight$_{j,t}$	Index for the level of property right protection within the province that year. It is a sub-index of marketization index compiled by Fan, Wang and Yu (2016). Our data spans from 2010 to 2014, for year 2015, we fill in the value for 2015 using data from the previous year.
HighGovInt$_{j,t}$	An indicator variable which takes on the value of one if the *GovIntervention*$_{j,t}$ is higher than the median and zero otherwise.
LowPptRight$_{j,t}$	An indicator variable which takes on the value of one if the *PptRight*$_{j,t}$ is lower than the median and zero otherwise.

Table 1: Firm-specific and Non-firm-specific Topics

Panel A - Number of general, industry and firm-specific topics:

	Number of topics
Total	370
Industry-specific	60
Remaining	310
Firm-specific	103
General	103

Panel B - The number of unique firm that each topic is relevant to:

	Mean	Median	Q1	Q3	Std.
Firm Specific	48.48	40.00	20.00	71.00	16.00
General	314.31	218.00	184.00	372.50	229.26
Industry Specific	123.20	121.00	108.00	141.00	25.20

Panel C - Average weight of each category of topics in each report:

	Mean	Median	Q1	Q3	Std.
$WFirmspecific_{i,j,t}$	44.93%	37.07%	17.93%	69.34%	32.64%
$WGeneral_{i,j,t}$	31.68%	23.78%	10.38%	50.14%	25.36%
$WIndustry_{i,j,t}$	10.83%	4.25%	1.64%	10.85%	16.86%

Panel D - Average number of topics in each report:

	Mean	Median	Std.	Q1	Q3
Firm Specific	3.22	3.00	2.00	4.00	1.95
General	3.76	4.00	2.00	5.00	2.17
Industry Specific	0.81	1.00	0.00	1.00	0.91

For each topic, we first identify whether that topic is relevant to a firm i. We assume that the topic is relevant to firm i, if it takes up more than 1% of any analyst reports issued for firm i. From the pool of 370 topics, for any topics that are relevant to 50% or more firms in any particular industry we classify them as an industry specific topic. After removing the industry specific topics, we are left with 310 topics, we rank these topics based on the number of firms that they are relevant to. We classify the bottom 1/3 as firm specific topics and the topic 1/3 as general topics.

Table 2: Summary Statistics

	Mean	Median	Q1	Q3	Std.
$BM_{j,t}$	0.48	0.38	0.23	0.71	0.35
$Size_{j,t}$	16.15	16.01	15.27	16.95	1.14
$Volume_{j,t}$	16.95	16.89	16.12	17.75	1.16
$Stdret_{j,t}$	3.06	2.69	2.25	3.33	1.36
$Horizon_{r,i,j,t}$	0.53	0.49	0.29	0.72	0.34
$Experience\_Firm_{i,j,t}$	1.13	0.50	0.00	1.85	1.43
$Experience_{r,i,t}$	2.91	2.81	1.39	4.44	1.75
$Brokersize\_Analyst_{i,t}$	0.10	0.10	0.06	0.13	0.05
$Brokersize\_Firms_{i,t}$	0.90	0.96	0.71	1.16	0.36
$Institutions\_share_{j,t}$	5.14	2.34	0.21	7.24	6.90
$Specialization_{i,t}$	3.18	3.00	2.00	4.00	2.12
$Star_{i,t}$	0.09	0.00	0.00	0.00	0.28
% Foreign $Analysts_{j,t}$	0.01	0.00	0.00	0.00	0.03
$Loss_{j,t}$	0.03	0.00	0.00	0.00	0.16
$Distance_{i,j,t}$	0.07	0.00	0.00	0.00	0.25
$Foreign\_Ownership_{j,t}$	0.29	0.00	0.00	0.00	1.04
$Synch_{j,t}$	-0.14	0.00	-0.50	0.20	0.66
$Fund\_Corr_{j,t}$	0.22	0.08	0.00	0.43	0.27
$Herf_{j,t}$	0.20	0.05	0.04	0.07	0.47
$STD\_ROA_{j,t}$	0.04	0.02	0.01	0.04	1.05
$NIND_{j,t}$	1059.12	1466.00	200.00	1746.00	766.61
$FirmSpecNews_{j,t}$	0.04	0.03	0.01	0.06	0.04
Political $Words_{i,j,t}$	0.31	0	0	0.20	2.39
Cosin_ $Similarity_{i,j,t}$	0.66	0.60	0.43	0.82	0.34
$LDA\_Similarity_{, i,j,t}$	-1.76	-1.78	-1.83	-1.71	0.09

Table 3: Firm-specific Information and Analyst Forecast Error (firm fixed effects)

	(1)	(2)	(3)	(4)	(5)
			Relative Forecast Error$_{i,j,t}$		
WFirmspecific$_{i,j,t}$	-0.076***			-0.072***	-0.072***
	(-5.81)			(-5.18)	(-5.21)
WNonFirmspecific$_{i,j,t}$		0.029**		0.021**	
		(2.34)		(2.21)	
WGeneral$_{i,j,t}$			0.026*		0.028*
			(1.83)		(1.82)
WIndustry$_{i,j,t}$			0.038*		0.036
			(1.70)		(1.42)
BM$_{j,t}$	0.049***	0.049***	0.049***	0.050***	0.050***
	(7.43)	(7.47)	(7.47)	(7.50)	(7.51)
Size$_{j,t}$	-0.026***	-0.025***	-0.025***	-0.026***	-0.026***
	(-12.05)	(-11.93)	(-11.93)	(-12.04)	(-12.04)
Volume$_{j,t}$	-0.008***	-0.008***	-0.008***	-0.007***	-0.007***
	(-3.30)	(-3.33)	(-3.33)	(-3.18)	(-3.19)
Stdret$_{j,t}$	0.003***	0.003**	0.003**	0.003***	0.003***
	(2.75)	(2.50)	(2.49)	(2.82)	(2.82)
Horizon$_{r,i,j,t}$	0.122***	0.122***	0.122***	0.122***	0.122***
	(26.18)	(26.15)	(26.15)	(26.15)	(26.15)
Experience_Firm$_{i,j,t}$	-0.008***	-0.008***	-0.008***	-0.009***	-0.008***
	(-4.55)	(-4.57)	(-4.57)	(-4.59)	(-4.58)
Experience$_{r,i,t}$	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***
	(-6.76)	(-6.66)	(-6.66)	(-6.72)	(-6.71)
Brokersize_Analyst$_{i,t}$	-0.014	-0.010	-0.010	-0.012	-0.012
	(-0.28)	(-0.21)	(-0.20)	(-0.25)	(-0.24)
Brokersize_Firms$_{i,t}$	-0.026***	-0.027***	-0.027***	-0.025***	-0.025***
	(-3.57)	(-3.73)	(-3.73)	(-3.48)	(-3.49)
Institutions_share$_{j,t}$	0.002***	0.002***	0.002***	0.002***	0.002***
	(5.72)	(5.82)	(5.82)	(5.72)	(5.71)
Specialization$_{i,t}$	-0.003***	-0.003***	-0.003***	-0.003***	-0.003***
	(-3.64)	(-3.68)	(-3.67)	(-3.58)	(-3.58)
Star$_{i,t}$	-0.001	-0.000	-0.000	-0.001	-0.001
	(-0.09)	(-0.03)	(-0.04)	(-0.08)	(-0.09)
% Foreign Analysts$_{j,t}$	0.064	0.055	0.055	0.060	0.060
	(1.23)	(1.06)	(1.06)	(1.14)	(1.15)
Loss$_{j,t}$	-0.025***	-0.025***	-0.025***	-0.025***	-0.025***
	(-3.85)	(-3.90)	(-3.91)	(-3.87)	(-3.87)
Distance$_{i,j,t}$	0.009	0.006	0.005	0.010	0.010
	(1.05)	(0.66)	(0.66)	(1.20)	(1.20)
Foreign_Ownership$_{j,t}$	0.003*	0.003*	0.003*	0.003*	0.003*
	(1.72)	(1.73)	(1.74)	(1.77)	(1.78)
Fixed Effects	Firm	Firm	Firm	Firm	Firm
Std. Cluster	Firm	Firm	Firm	Firm	Firm
Observations	87332	87332	87332	87332	87332

Adjusted R^2	0.075	0.074	0.074	0.075	0.075

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The above table include 87,332 observations, each observation represents the latest report by analyst i issued for firm j for year t. Relative forecast error$_{i,j,t}$ is the proportional mean absolute forecast error relating to that particular report. More specifically, it is the ratio of the difference between the absolute forecast error of analyst *i* forecasting for firm *j*'s fiscal year *t* earnings and the average absolute forecast error across all analyst forecasts of firm *j*'s fiscal year *t* earnings, to the mean absolute forecast error. i.e. $\frac{AFE_{i,j,t}-AvgAFE_{j,t}}{AvgAFE_{j,t}}$. $WFirmspecific_{i,j,t}$ represents the sum of all firm specific topic weight, $WGeneral_{i,j,t}$ represents the sum of all general topic weight. $WIndustry_{i,j,t}$ represents the sum of all industry-specific topic weight. $WNonFirmspecific_{i,j,t}$ is the sum of $WGeneral_{i,j,t}$ and $WIndustry_{i,j,t}$.

Table 4: Validation for our proxy of firm-specific information

	(1)	(2)	(3)	(4)	(5)	(6)
		$\text{Synch}_{j,t}$			$\text{FirmSpecificEarnings}_{j,t}$	
$Avg\_Firmspecific_{j,t}$	-0.145**			0.010**		
	(-3.10)			(3.41)		
$Avg\_General_{j,t}$		0.151*			-0.004	
		(2.12)			(-1.40)	
$Avg\_Industry_{j,t}$			0.294*			-0.001
			(2.26)			(-0.23)
$\text{Fund\_Corr}_{j,t}$	0.018	0.017	0.019	-0.013***	-0.013***	-0.013***
	(0.34)	(0.33)	(0.38)	(-4.22)	(-4.12)	(-4.13)
$\text{In(Herf}_{j,t})$	0.014	0.014	0.014	0.005**	0.005**	0.005**
	(0.87)	(0.83)	(0.82)	(2.89)	(2.85)	(2.86)
$\text{STD\_ROA}_{j,t}$	-0.080**	-0.081**	-0.082**	0.002	0.002	0.002
	(-3.02)	(-3.10)	(-3.04)	(1.72)	(1.71)	(1.86)
$\text{In(NIND}_{j,t})$	0.019	0.022	0.018	0.005	0.005	0.005
	(0.24)	(0.27)	(0.23)	(1.44)	(1.39)	(1.42)
$\text{Size}_{j,t}$	-0.380***	-0.384***	-0.386***	0.017***	0.018***	0.018***
	(-6.82)	(-6.94)	(-6.99)	(6.11)	(6.13)	(6.15)
Fixed Effects	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year
Std. Cluster	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year
Observations	10384	10384	10384	10384	10384	10384
Adjusted R^2	0.336	0.336	0.336	0.573	0.572	0.572

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The above table include 10,384 firm-year observations. $\text{Synch}_{j,t}$ is a firm-specific measures of return synchronicity using the methodology outlined in Piotroski and Roulstone (2004). $\text{FirmSpecificEarnings}_{j,t}$ is the firm-specific component of earnings innovation estimated using the procedure outlined in Ball and Brown (1967). $Avg\_Firmspecific_{j,t}$ is the average firm specific topic weights across all reports issued for firm j in year t. $Avg\_General_{j,t}$ is the average general topic weights across all reports issued for firm j in year t. $Avg\_Industry_{j,t}$ is the average industry specific topic weights across all reports issued for firm j in year t.

Table 5: Political influence at the firm-level and firm-specific information

	(1)	(2)	(3)
	Relative Forecast Error$_{i,j,t}$		
$WFirmspecific_{i,j,t}$×$HighPolitical1_{j,t}$	-0.039***		
	(-2.95)		
$HighPolitical1_{i,j,t}$	-0.003		
	(-0.94)		
$WFirmspecific_{i,j,t}$× $HighPolitical2_{j,t}$		-0.031**	
		(-2.56)	
$HighPolitical2_{i,j,t}$		-0.004	
		(-1.17)	
$WFirmspecific_{i,j,t}$× $HighPolitical3_{j,t}$			-0.112***
			(-5.93)
$HighPolitical3_{i,j,t}$			-0.003
			(-1.16)
$WFirmspecific_{i,j,t}$	-0.071***	-0.079***	0.028*
	(-4.63)	(-5.92)	(1.95)
BM$_{j,t}$	0.048***	0.048***	0.045***
	(7.29)	(7.30)	(6.82)
Size$_{j,t}$	-0.026***	-0.026***	-0.029***
	(-12.08)	(-12.01)	(-13.43)
Volume$_{j,t}$	-0.008***	-0.008***	-0.006***
	(-3.37)	(-3.67)	(-2.69)
Stdret$_{j,t}$	0.003***	0.003**	0.003**
	(2.61)	(2.20)	(2.04)
Horizon$_{r,i,j,t}$	0.122***	0.123***	0.122***
	(26.19)	(26.21)	(26.15)
Experience_Firm$_{i,j,t}$	-0.008***	-0.008***	-0.008***
	(-4.40)	(-4.37)	(-4.12)
Experience$_{r,i,t}$	-0.009***	-0.010***	-0.009***
	(-6.59)	(-6.96)	(-6.50)
Brokersize_Analyst$_{i,t}$	-0.009	-0.017	-0.001
	(-0.19)	(-0.36)	(-0.02)
Brokersize_Firms$_{i,t}$	-0.027***	-0.026***	-0.028***
	(-3.72)	(-3.66)	(-3.95)
Institutions_share$_{j,t}$	0.002***	0.002***	0.002***
	(5.68)	(5.65)	(5.37)
Specialization$_{i,t}$	-0.003***	-0.003***	-0.003***
	(-3.70)	(-3.65)	(-3.79)
Star$_{i,t}$	-0.001	-0.001	-0.001
	(-0.09)	(-0.13)	(-0.12)
% Foreign Analysts$_{j,t}$	0.065	0.067	0.079
	(1.24)	(1.28)	(1.49)
Loss$_{j,t}$	-0.026***	-0.026***	-0.023***
	(-3.87)	(-4.00)	(-3.33)
Distance$_{i,j,t}$	0.009	0.008	0.006

	(1.03)	(1.00)	(0.68)
Foreign_Ownership$_{j,t}$	0.003*	0.003*	0.003
	(1.67)	(1.73)	(1.60)
Fixed Effects	Firm	Firm	Firm
Std. Cluster	Firm	Firm	Firm
Observations	87332	87332	87332
Adjusted R^2	0.075	0.076	0.082

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The above table include 87,332 observations, each observation represents the latest report by analyst i issued for firm j for year t. *Relative forecast error*$_{i,j,t}$ is the proportional mean absolute forecast error relating to that particular report. More specifically, it is the ratio of the difference between the absolute forecast error of analyst *i* forecasting for firm *j*'s fiscal year *t* earnings and the average absolute forecast error across all analyst forecasts of firm *j*'s fiscal year *t* earnings, to the mean absolute forecast error. i.e. $\frac{AFE_{i,j,t} - AvgAFE_{j,t}}{AvgAFE_{j,t}}$. *WFirmspecific*$_{i,j,t}$ represents the sum of all firm specific topic weight. *HighPolitical1*$_{j,t}$ is an indicator variable which takes the value of 1 when the average *Cosin_ Similarity*$_{i,j,t}$ across all of firm j's analyst reports is higher than the median. *HighPolitical2*$_{j,t}$ is an indicator variables which takes on the value of 1 when the average $LDASimilarity$ across all of firm j's analyst reports is higher than median. *HighPolitical3*$_{j,t}$ is an indicator variable which takes on the value of 1 when the average *Political Words*$_{i,j,t}$ across all of firm j's analyst reports is higher than the median.

Table 6: Political influence at the state-level and firm-specific information

	(1)	(2)	(3)	(4)
	Relative Forecast Accuracy$_{i,j,t}$			
$WFirmspecific_{i,j,t} \times Turnover_{j,t}$	-0.038***			
	(-4.00)			
$Turnover_{j,t}$	0.009**			
	(2.00)			
$WFirmspecific_{i,j,t} \times Policy_{j,t}$		-0.101***		
		(-3.92)		
$Policy_{j,t}$		0.049***		
		(2.83)		
$WFirmspecific_{i,j,t} \times HighGovInt_{j,t}$			-0.057***	
			(-2.99)	
$HighGovInt_{j,t}$			-0.021	
			(-1.61)	
$WFirmspecific_{i,j,t} \times LowPptRight_{j,t}$				-0.055***
				(-2.75)
$LowPptRight_{j,t}$				-0.010
				(-1.57)
$WFirmspecific_{i,j,t}$	-0.072***	-0.024	-0.064***	-0.061***
	(-5.45)	(-1.12)	(-4.21)	(-3.59)
BM$_{j,t}$	0.046***	0.049***	0.050***	0.051***
	(6.90)	(7.44)	(7.38)	(7.31)
Size$_{j,t}$	-0.026***	-0.026***	-0.026***	-0.027***
	(-11.91)	(-11.92)	(-11.97)	(-11.63)
Volume$_{j,t}$	-0.009***	-0.007***	-0.009***	-0.010***
	(-3.74)	(-3.21)	(-3.86)	(-4.00)
Stdret$_{j,t}$	0.004***	0.003***	0.003**	0.003**
	(3.33)	(2.76)	(2.23)	(2.57)
Horizon$_{r,i,j,t}$	0.122***	0.122***	0.123***	0.121***
	(26.18)	(26.17)	(26.27)	(26.02)
Experience_Firm$_{i,j,t}$	-0.008***	-0.008***	-0.008***	-0.009***
	(-4.54)	(-4.49)	(-4.47)	(-4.59)
Experience$_{r,i,t}$	-0.010***	-0.010***	-0.008***	-0.013***
	(-6.80)	(-6.83)	(-5.62)	(-8.60)
Brokersize_Analyst$_{i,t}$	-0.019	-0.013	-0.028	-0.019
	(-0.40)	(-0.26)	(-0.57)	(-0.38)
Brokersize_Firms$_{i,t}$	-0.025***	-0.026***	-0.023***	-0.023***
	(-3.43)	(-3.68)	(-3.26)	(-3.22)
Institutions_share$_{j,t}$	0.002***	0.002***	0.002***	0.002***
	(5.75)	(5.73)	(5.63)	(5.88)
Specialization$_{i,t}$	-0.003***	-0.003***	-0.003***	-0.003***
	(-3.53)	(-3.63)	(-3.59)	(-3.66)
Star$_{i,t}$	-0.001	-0.001	-0.001	-0.001
	(-0.13)	(-0.10)	(-0.13)	(-0.12)
% Foreign Analysts$_{j,t}$	0.073	0.066	0.091*	0.047

	(1.38)	(1.26)	(1.66)	(0.81)
$Loss_{j,t}$	-0.025***	-0.025***	-0.026***	-0.029***
	(-3.71)	(-3.77)	(-3.86)	(-4.07)
$Foreign\_Ownership_{j,t}$	0.003*	0.003*	0.003	0.003
	(1.69)	(1.74)	(1.52)	(1.56)
Fixed Effects	Firm	Firm	Firm	Firm
Std. Cluster	Firm	Firm	Firm	Firm
Observations	87332	87332	87332	87332
Adjusted R^2	0.075	0.075	0.075	0.076

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The above table include 87,332 observations, each observation represents the latest report by analyst i issued for firm j for year t. *Relative forecast error*$_{i,j,t}$ is the proportional mean absolute forecast error relating to that particular report. More specifically, it is the ratio of the difference between the absolute forecast error of analyst *i* forecasting for firm *j*'s fiscal year *t* earnings and the average absolute forecast error across all analyst forecasts of firm *j*'s fiscal year *t* earnings, to the mean absolute forecast error. i.e. $\frac{AFE_{i,j,t}-AvgAFE_{j,t}}{AvgAFE_{j,t}}$. *Turnover*$_{j,t}$ is an indicator variable which takes on the value of 1 if firm j's headquarter province experienced a key politician turnover in year t or t-1. *Policy*$_{j,t}$ is an indicator variable which takes on the value of 1 if the industry in which firm j belongs to is being promoted by firm j's headquarter province in year t. *GovIntervention*$_{j,t}$ is an index for the level of province level government intervention within the province in which company j is headquartered in, higher value indicate more intervention. *Pptright*$_{j,t}$ is an index for the level private property right within the provicince in which company j is headquartered in, higher value indicate better property right protection. *HighGovInt*$_{j,t}$ is an indicator variable which takes the value of 1 if the *GovIntervention*$_{j,t}$ is higher than the median. *LowPptRight*$_{j,t}$ is an indicator variable which takes the value of 1 if the *Pptright*$_{j,t}$ is higher than the median.

Table 7: Additional Tests - Are analyst providing firm-specific information to the market?

Panel A – Analyst turnover frequency

	n	%
Unique # analysts in our sample	3175	100%
Analyst that turnover during our sample period	1068	34%
Analyst that turnover in 2011	*298*	
Analyst that turnover in 2012	*322*	
Analyst that turnover in 2013	*448*	

Panel B – Firm specific analyst vs. non-firm specific analyst characteristics

	Mean		Median	
	$Firmspecific_{i,j,t}$	# Firms	$Firmspecific_{i,j,t}$	# Firms
Firm Specific Analysts	0.78	20.78	0.76	10.00
Non-Firm Specific Analysts	0.23	27.05	0.25	14.00
t-stat or z-score	*36.62****	*2.07***	*30.59****	*1.51**

	(1)	(2)
	$Synch_{j,t}$	$FirmSpecificEarnings_{j,t}$
$Departure_{i,t}$×$FirmSpecAnalyst_{i}$	0.334^{**}	0.001
	(2.47)	(0.46)
$FirmSpecAnalyst_{i}$	-0.031^{*}	-0.000
	(-2.34)	(-0.10)
$Departure_{i,t}$	0.009	-0.000
	(0.14)	(-0.30)
$Fund\_Corr_{j,t}$	0.196^{**}	-0.001
	(2.57)	(-0.17)
In($Herf_{j,t}$)	-0.008	0.005
	(-0.07)	(1.78)
$STD\_ROA_{j,t}$	-0.132	0.374^{***}
	(-0.68)	(11.92)
In($NIND_{j,t}$)	-0.176	0.005
	(-0.65)	(0.76)
$Size_{j,t}$	-0.303^{***}	0.021^{***}
	(-4.91)	(8.79)
Fixed Effects	Firm, Year	
Std. Cluster	Firm, Year	
Observations	13685	13685
Adjusted R^2	0.508	0.595

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The table above includes 13,685 firm-year observations. The unit of analysis is firms with a departing analyst (*i*)- pre-/post-(*t*). Only departure of firm-specific and non-firm-specific analysts are included. $Synch_{j,t}$ is a firm-specific measures of return synchronicity using the methodology outlined in Piotroski and Roulstone (2004). $FirmSpecificEarnings_{j,t}$ is the firm-specific component of earnings innovation estimated using the procedure outlined in Ball and Brown (1967). $Departure_{i,t}$ is an indicator variable which takes a value of 1 for observations after the departure event. $FirmSpecAnalyst_{i}$ is an indicator variable which takes the value of 1 if the departing analyst's $WFirmspecific_{i,j,t}$ is ranked in the top quartile.