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

Forester Wong*

T.J. Wong*

Tianyu Zhang#

November 2018

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 Please do not cite or circulate without the authors' permission.

This paper examines whether and how politics shapes the kind of information that enhances analysts' forecast accuracy in a relational economy. Since political influence is exerted on firms primarily through relationships, information about firms' performance is highly specific. Even for firms that are within the same industry, these relationships can differ significantly. We posit that politics increases the idiosyncrasy of analysts' information that is accuracy-enhancing. 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 when political influence on firms increases, the idiosyncratic topics (i.e. topics that are specific to fewer firms) in the analysts' reports are more positively associated with their relative forecast accuracy. However, we do not find that politics influences the relation between industry-specific topics (i.e. topics that are specific to firms in the same industry) and forecast accuracy. 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 and workshop participants at USC for their helpful 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 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 whether and how politics determines the kind of information that enhances analysts' forecast accuracy. We predict that when politics becomes more influential on the future performance of a firm (e.g. due to changes in political connections or government contracts or policies), the information that is important for analysts to forecast earnings becomes even more idiosyncratic since the political influence is relational and highly specific.[2]

However, ex-ante, it is not clear whether the accuracy-enhancing information is more idiosyncratic when the political influence on these firms heightens. On the one hand, financial analysts will develop their core competence through analyzing industry or market trends, since information about the firms' relationships is not readily available in the market. This is because firms are less inclined to provide voluntary disclosures about their relational contracts or political connections than arm's length contracts because relationships are often proprietary and hard to verify (Fan and Wong, 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

[1] Examples of how relationships are essential for evaluating firms' contracts and making forecasts are presented in the Appendix I.
[2] We focus on idiosyncrasy of information rather than measuring directly information about political relationships because analysts will not discuss such proprietary and sensitive relationships in their reports.

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 less idiosyncratic. Thus, in a heightened political environment, analysts are even more likely to specialize in gathering information that is more industry-specific and less idiosyncratic in nature.

On the other hand, financial analysts will try to establish comparative advantage by gathering information about the firms' relationships that are highly idiosyncratic and essential for evaluating the firms' contracts.[3] Although information pertains to the firms' relationships is in short supply, it is imperative for generating earnings forecasts. 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. Thus, when the political influence on firms increases, analysts will gather information that is more idiosyncratic to enhance forecast accuracy.

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. Further, the significant cross-sectional differences in policies between provinces and industries provide a setting to examine the impact that different levels of political influence can have on the information environment within one country.

[3] 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).

To measure the kind of information that can enhance Chinese financial analysts forecast accuracy, we examine the content of the analyst reports. We posit that the textual information contained in the analysts' report should reflect the information being used by them in making 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 or less idiosyncratic information. In particular, we use Latent Dirichlet Allocation (LDA hereafter), a topical modeling method, to identify 370 topics in the financial analysts' reports.[4] Using a set of objective criteria, we then classify these topics into idiosyncratic topics and industry-specific topics.

Our classification is done based on how these topics are distributed across firms. First, we isolate topics that likely contain industry-specific information. 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.[5] We consider that 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 terciles. We classify the bottom tercile, which are topics mentioned in fewer firms, as idiosyncratic 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 idiosyncratic information. Similarly, for topics that are concentrated within an industry, we expect that the

[4] 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.
[5] The 1-digit SIC classification allows us to sharpen the power of our test by identifying *fewer* topics that are specific to larger industry groups. Our results are robust to using the 2-digit SIC classification (see Subsection 3.3).

content of these topics is likely to be at the industry level, which is therefore less idiosyncratic as well. As will be discussed in Subsection 4.5.2, we find qualitatively similar results in our main tests when we use different cutoffs for the idiosyncratic and industry-specific topic classifications.

Using these measures of analysts' information, we first examine the baseline relation between analysts' relative forecast accuracy and their information idiosyncrasy. 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. We find that idiosyncratic topics are significantly positively associated with forecast accuracy. However, we find that industry-specific topics are significantly *negatively associated with* analysts' forecast accuracy.[6] 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.

Next, we test whether the positive association between idiosyncratic topics and analysts' relative forecasts accuracy will increase when the underlying firms is experiencing greater political influence. We use three approaches to measure the heightening of the political influence. First, we proxy for the level of political information using the average amount of political content in a firm's analyst reports. We measure the political content of analyst reports using a count of political words or the analyst reports' similarity in thematic content with the People's Daily, a Party newspaper

[6] One interpretation of this result is that by focusing too much on 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 industry trends when in fact the firm's close relationship to an incoming politician may cause this firm to deviate from industry trends in future years.

that promotes government policies. We find that the positive association between idiosyncratic topics 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 idiosyncratic topics and analysts' forecast accuracy is stronger in the year before and the year of 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.[7] When provincial leaders are promoting certain industries, the firms belonging to those industries will experience increased political influence. For these set of firms, we document that the relation between idiosyncratic topics and analysts' forecast accuracy is stronger during the years in which the underlying industry is being supported.

Third, we interact idiosyncratic information with two separate measures of the institutional environment of the province in which the firms operate. We find that idiosyncratic topics are even more valuable to forecasting when the firms operate in provinces with greater government intervention or weak private property rights protection. Also, none of the three measures of political increases the association between industry-specific topics and forecast accuracy. In fact, the industry-specific topics are *more* significantly *negatively* associated with forecast accuracy in

[7] 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.

periods with politician turnover, and in provinces with more government intervention or weak property rights protection. These results provide supporting evidence 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 idiosyncratic topics of a firm is significantly positively associated with the level of its idiosyncratic component of the earnings innovation. However, the average relative weight of the industry-specific topics is not associated with the idiosyncratic 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 idiosyncratic topics in their analyst reports are associated with lower stock return synchronicity. In contrast, firms with higher average relative weight of industry-specific topics are associated with higher 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 idiosyncratic information in the market. By showing that the idiosyncratic topics of the analyst reports can impact stock returns and increase the market's information environment, it provides credence to our main argument that the idiosyncratic 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 idiosyncratic topics prior to its departure vs. those that produce fewer idiosyncratic 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 idiosyncratic topics than those that produce fewer idiosyncratic 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 idiosyncratic 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 idiosyncratic information, especially in environment where the political influence is heightened.

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 idiosyncratic and it provides them with a stronger comparative advantage in forecasting earnings. 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 how politics affect the kind of topics in the financial analyst reports that are associated with greater forecast accuracy.

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 more (less) idiosyncratic 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 idiosyncratic information and firms' stock return synchronicity provides support to their interpretation that low synchronicity is driven by idiosyncratic information but not noise. This interpretation is further supported by the result that the idiosyncratic information used in the study is associated with higher forecast accuracy.[8]

[8] 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 idiosyncratic information to the market, thereby narrowing the gap in stock return synchronicity of relationship-based economies with market-based economies.

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 and Research Design

Our research question is to examine if politics increases or decreases the importance of the specificity of financial analysts' information in making earnings forecasts. We posit that when the political influence on firms increases, idiosyncratic information that captures 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 idiosyncratic because the government exerts influence on firms via relationships and these relationships are highly idiosyncratic.

However, it is unclear if Chinese analysts will even gather idiosyncratic information to make their forecasts in the first place because this type of information is typically not publicly available. In a relational economy such as the one in China, firms often turn to informal mechanisms for contracting since laws and courts are ineffective against opportunism. These relational contracts increase firms' opacity because relationships, especially those with the politicians, 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 idiosyncratic 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 idiosyncratic information.

In such an information environment, Chinese financial analysts cannot rely on publicly available idiosyncratic information for their analysis and formulating earnings forecasts. Even in the U.S. where there is more idiosyncratic 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 primarily provide industry or general information rather than idiosyncratic 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 idiosyncratic information to the market.

On the other hand, financial analysts may develop their comparative advantage by providing idiosyncratic 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.[9] With this knowledge, however, the analysts can combine it with a seemingly immaterial

[9] China's Reg FD stated in China Securities Regulatory Commission Articles 40 and 128.

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.[10] Since every firm's relationship set is unique, this kind of information is likely to be idiosyncratic. Further, when there is a heightening of political influence, idiosyncratic information becomes even more important for analysts to make their forecasts. 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.[11] We also use the similarity in thematic content between the analyst report and the political and economic sections

[10] 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.

[11] 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.

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 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 idiosyncratic 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. Our hypothesis is as follows:

The positive association between idiosyncratic information and relative earnings forecast accuracy increases as the political influence on the firms increases.

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.

Finally, the data for the various proxies of political influence are collected from a number of sources. The People's Daily articles are obtained from People.com website.[12] Data on political promotion events are hand-collected by searching information published in the "Chinese Personnel Database" and "China VIPs" from China Information Bank and supplemented by Google web searches. The industrial policy data are obtained from the various websites of the provincial governments. The data on government intervention and property rights protection of the provinces are obtained from Fan et al. (2014).

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.

[12] http://www.people.com.cn/GB/43063/43070/index.html.

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.[13] 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.[14] We have also set our topics to 200 and 500, and as discussed in Subsection 4.5.2, the results of the main test remain qualitatively the same.

3.3 Topic classification

We identify idiosyncratic topics and industry-specific topics based on how each topic is distributed across firms. We begin by isolating industry-specific topics. 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. Our main results are robust to 2-digit SIC industry and also alternative cut-offs (40% and 60%). We consider that 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. As presented in Table 1 Panel

[13] 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 idiosyncratic 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.

[14] 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.

A, this procedure yields 60 industry-specific topics. For the remaining 310 topics, we divide them into terciles. We classify the bottom tercile, which are topics mentioned in fewer firms, as idiosyncratic topics. For illustrative purposes, we also include in Table 1 the summary statistics of the other unclassified topics. On average the idiosyncratic topics are relevant to ~78 firms, while the industry topics are relevant to ~123 firms and the idiosyncratic topics are~ relevant to 233 firms (Panel B). While on average idiosyncratic topics account for ~45% of the analyst reports, and the industry topics account for ~11% of the analyst reports and the idiosyncratic topics accounts for ~44% of the analyst reports. Appendix II (III) includes representative paragraphs from five analyst reports that contain the idiosyncratic (industry-specific) topics.

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 idiosyncratic and industry-specific information. Finally, using three different sets of proxies for political influence, we show that the positive association between idiosyncratic information and relative earnings forecast accuracy is significantly stronger when the political influence on the firms increases.

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.,

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

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:

$$\textit{Relative Forecast Error}_{i,j,t} = \alpha + \beta\ \textit{WIdiosyncratic}_{i,j,t} + \text{Control Variables} + \ FE_{Firm,} \tag{2}$$

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, *WIdiosyncratic*$_{i,j,t}$, is the relative weight of idiosyncratic topics in the analyst report. A negative and significant coefficient on *WIdiosyncratic*$_{i,j,t}$ will indicate that idiosyncratic 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.[15]

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

[15] 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.

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 IV. 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 $WIdiosyncratic_{i,j,t}$ is significantly negative, consistent with our conjecture that, analysts with more idiosyncratic information have a higher comparative advantage. However, moving onto industry-specific information, the coefficient on $WIndustry_{i,j,t}$ is positive and significantly different from zero, implying that having industry-specific information does not increase an analyst's comparative advantage. Our results remain similar, when we include analyst fixed effects in Column 4 and analysts and year fixed effects in Column 5.

4.3 Topic validation

We use two different methods that prior literature employs to capture idiosyncratic information to validate our topical classifications. First, similar to Piotroski and Roulstone (2004), we use stock return synchronicity to measure the amount of idiosyncratic, 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:

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

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 idiosyncratic information. To test whether our classification reflects idiosyncratic and non-idiosyncratic information, we estimate the following pooled firm-year model:

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

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 idiosyncratic topic weight ($Avg\_Idiosyncratic_{j,t}$), and the average industry-specific topic weight ($Avg\_Industry_{j,t}$) 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 $Avg\_Idiosyncratic_{j,t}$ is high, firm *j*'s stock return synchronicity is significantly lower (column 1), implying that more idiosyncratic information is impounded into firm *j*'s stock prices. In contrast, when $Avg\_Industry_{j,t}$ is high (column 2), firm *j*'s stock return synchronicity is also significantly higher, implying that less idiosyncratic information is impounded into firm *j*'s stock prices.

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 idiosyncratic earnings component. More specifically, for each firm-year observation, we regress the change in ROA ($\Delta\mathrm{ROA}_{j,t-\delta}$) on the change in quarterly market ROA ($\Delta\mathrm{ROA}_{mk,t-\delta}$) and industry change in ROA ($\Delta\mathrm{ROA}_{ind,t-\delta}$), or:

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

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:

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

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

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

Higher value of $IdiosyncraticEarnings_{j,t}$ indicates more idiosyncratic earnings innovation. To test whether our classification reflects idiosyncratic information, we replace $SYNCH_{j,t}$ in equation (5) with $IdiosyncraticEarnings_{j,t}$:

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

As demonstrated in column 3, the coefficient on $Avg\_Idiosyncratic_{j,t}$ is positive and significant, implying that the average idiosyncratic topical weight is positively associated with the amount of idiosyncratic earnings innovation. In contrast, the coefficient on $Avg\_Industry_{j,t}$ is insignificantly different from zero (columns 4), indicating that it does not contain any idiosyncratic information.

4.4 Does political influence alter the relationship between idiosyncratic 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.[16] 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.

[16] 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.

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:

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

, 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 Panel A, *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 *WIdiosyncratic*$_{i,j,t}$ to examine the importance of idiosyncratic information for firms with high level of political influence. In Columns (1) to (3), we find that the relative forecast error is significantly lower when the weight of both idiosyncratic topics and political content is higher. This support our hypothesis that political influence increases the importance of idiosyncratic information.

In Table 5 Panel B, we examine the impact of political influence on industry-specific information by replacing $WIdiosyncratic_{i,j,t}$ with $WIndustry_{i,j,t}$. Across our three textual proxies of political influence, we find no evidence that political influence alters the importance of industry 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 influence. In Table 6 Panel A, $Turnover_{j,t}$ 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 $WIdiosyncratic_{i,j,t}$ with $Turnover_{j,t}$ and in Column (2) when we interact $WIdiosyncratic_{i,j,t}$ with $Policy_{j,t}$, we find that both interaction terms are negative and significant, indicating that for these firms, idiosyncratic information is more important during these periods with heightened political influence. Overall, our results suggest that the relationship between idiosyncratic 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 idiosyncratic content to play an even more significant role in the analyst's comparative advantage in forecasting earnings.

Table 6 Panel A columns 3 and 4 show that idiosyncratic 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, idiosyncratic information is providing an even stronger comparative advantage to the financial analysts to forecast.

In Table 6 Panel B, we replace the $WIdiosyncratic_{i,j,t}$ with $WIndustry_{i,j,t}$, across our four proxies of heightened political pressures, we find that three of them are positive and significant. This is consistent with our conjecture the idiosyncrasy of relationships may outweigh the common industry-level components in shaping the type of information analysts need for forecasting.

4.5 Additional Tests

4.5.1 Do Chinese analysts provide idiosyncratic information to the market?

Our earlier test in Table 4 is consistent with our conjecture that analysts are providing idiosyncratic 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 idiosyncratic content when there is more idiosyncratic 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 one years before departure. As shown in Table 7 Panel A, during our sample period, there is a total 1,128 analyst departures, and these departures are evenly distributed across the four years from 2010 to 2013.[17] 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 idiosyncratic or non-idiosyncratic analysts. To do that, we first rank all the analysts based on the amount of idiosyncratic information in all their reports. We then classify an analyst to be an idiosyncratic (non-idiosyncratic) 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 an idiosyncratic analyst's departure has on post-departure stock return synchronicity, relative to that of the departure of a non-idiosyncratic analyst. The test is akin to a difference-in-differences model where we use the non-idiosyncratic analysts' departure to control for unobserved factors that may confound the departing sample. The departure of idiosyncratic analysts is the treatment group. Equation (12) shows the regression model for the departure test:

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

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

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

to within one year of the analyst's departure. Also, to ensure that departures are not affected by other confounding events, we require firms to have no other departing analysts 120 days before this event. 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.

$IdiosyncraticAnalyst_i$ is an indicator variable that takes a value of 1 if departure event involves an idiosyncratic-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 $Departure_{i,t} \times IdiosyncraticAnalyst_i$ is our main variable of interest; it captures the incremental effect of an analyst's departure when the departure involves an idiosyncratic-analyst. We predict that the increase in synchronicity following an idiosyncratic-analyst's departure will be greater than that of the non-idiosyncratic 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 IV.

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 idiosyncratic topics than those that produce fewer idiosyncratic topics. This suggests that our earlier finding that Chinese analysts are providing idiosyncratic 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 $Departure_{i,t} \times IdiosyncraticAnalyst_i$

and $Departure_{i,t}$ are both insignificantly different from zero.[18] 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 idiosyncratic and industry-specific topics. For idiosyncratic topics we use the bottom 50, 77 (quartiles) and 120 topics. For industry-specific topics, we alter the cut-offs to topics mentioned in 40% of firms and 60% of firms in any industries. We repeat the regressions in Table 3 using these new classifications for idiosyncratic and industry-specific topics. Our results show that the coefficient on $WIdiosyncratic_{i,j,t}$ 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 bottom 50 topics as partition.

Further, as a placebo test, we re-run the regressions in Table 3 using the top 50, 77 (quartiles) and 120 topics, those that are least likely to be idiosyncratic. None of the coefficients on these topic weights are significantly different from zero using the three partitions for non-idiosyncratic 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 $WIdiosyncratic_{i,j,t}$ to be negative and significant at the one percent level. However, the coefficient on $WIndustry_{i,j,t}$ is not different from zero when 200 or 500 is used as the number of classification.

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

5. Conclusion

This paper finds that in a relationship-based economy such as China, it is the idiosyncratic information in the financial analysts' reports that give them the comparative advantage in forecasting earnings than non-idiosyncratic information. More importantly, we find that the level of political influence on a firm increases the positive association between idiosyncratic information and forecast accuracy. That is, when politics has a stronger influence on a firm, analysts will need more idiosyncratic information about the relationship between the politician and firm, to make their earnings forecasts. Our topical measures of idiosyncratic and industry-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, G., Wang, X. and Zhu, H., 2014. The marketization index of China: The process of regional marketization report 2013. Beijing: Press of Economic Science [in Chinese].

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 and the firms' stakeholders. 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 with a deep understanding of 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, an analyst with deep understanding of the companies would know that there was tension in the relationship between the owners of the target and the acquiring firms. 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 informed 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 informed 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, an informed 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 deep understanding of the brokerage house's nexus of connections 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: Representative paragraphs from analyst reports with idiosyncratic topics

Idiosyncratic Example 1: Travel Business 2012

索道项目推进顺利，超出市场预期。（1）索道运力是武陵源景区的最大瓶颈，杨家界索道是解决接待瓶颈的最好的手 段，也是政府最迫切推进的项目。但索道项目受制于交通条件，在国家宏观调控逐步深入的背景下，市场对于交通能 否在 2013 年得到解决存有疑虑，也对索道 2013 年能否投入运营存有疑虑。（2）虽然通过打通黄石寨-杨家界来解决 索道交通问题，效果没有"直通高速"那么好，但确保了索道 2013 年旺季之前能够投入运营，我们认为将超出了市场预期。

The ropeway project progressed smoothly and exceeded market expectations. (1) Cableway capacity is the biggest bottleneck in Wulingyuan Scenic Area. Yangjiajie Cableway is the best means to solve the bottleneck of reception, and ***it is also the most urgently promoted project of the (local) government.*** However, the ropeway project is subject to traffic conditions. Given the ***central government policy on corruption, the market has doubts*** about whether the traffic can be resolved in 2013, and there are doubts about whether the ropeway can be put into operation in 2013. (2) Although cableway was not as effective as "straight-through high-speed" way, but it ensured that the ropeway could be put into operation before the peak season in 2013, and we believe this will allow the company to exceed market expectations.

Idiosyncratic Example 2: Farming 2010

公司前身为_________，于 2001 年成立。董事长李成荃为我国杂交水稻国际开发专家组成员、农业 部杂交水稻专家顾问组顾问成员，在安徽省农科院作物所从事水稻研究 50 余年，其中杂交水稻育种 30 余年。因此公 司拥有丰富的专业研发人员资源和育种资源，这是种子行业关键制胜点之一，也是获得持续竞争力的条件之一。公司 目前致力于研发新品种，上半年又新增 4 项新品种。不断推陈出新，才能获得客户的信任，占领市场。

The company was formerly known as _______ and was established in 2001. ***Chairman Li Chengwei is a member of the Expert Group on Hybrid Rice International Development of China and a consultant member of the Hybrid Rice Expert Advisory Group of the Ministry of Agriculture.*** He has been engaged in rice research for more than 50 years in the Crop Research Institute of Anhui Academy of Agricultural Sciences, including hybrid rice breeding for more than 30 years. Therefore, the company has a wealth of professional R&D personnel resources and breeding resources. This is one of the ***key winning points*** in the seed industry and one of the conditions for ***sustained competitiveness***. The company is currently working on new varieties and has added four new varieties in the first half of the year. Constantly innovating in order to gain the trust of customers and occupy the market.

Idiosyncratic Example 3: Entertainment 2013

2013 年上半年电视剧业务"保质增量"稳步推进，电视台客户质量整体提升 1、2013 年上半年 4 部剧取得发行许可，7 部剧已经开拍/完成拍摄，尚未取得发行 许可。（1）上半年开播的电视剧有：《天真遇到现实》在东方、江苏、深圳三卫视首 播，获得全国卫视同时段收视冠军；徐纪周导演的《战雷》在北京、安徽、深圳、 云南四家卫视首轮播出，收视率略低于预期。（2）报告期内 4 部剧目取得发行许可 证：《风雷动》、《零炮楼》、《战雷》、《浪漫我心》。（ 3）已开机或者基本完成拍摄， 尚未取得发行许可证的剧目有 7 部：《金战尖兵》、《岁月如金》（原名《北京一家人》）、《留守男人》、《咱们结婚吧》（将于 2013 年在 CCTV -1 和湖南卫视黄金时段播出）、《美丽的契约》、《煮妇也疯狂》、《爱的合同》。其中，三部剧投资权益比超 90%；三 部剧权益比 50%～60%；一部剧权益比为 30%。 2、前五大剧目贡献核心收入，电视台客户质量整体提升。（1）上半年公司主营业务 收入前五名的影视剧分别为《战雷》、《零炮楼》、《永不磨灭的番号》、《歧路兄弟》、《媳妇的美好时代》，合计金额占公司同期主营业务收入的 94.33%。（2）从客户情 况来看，公司电视台客户质量整体提升。上半年公司五大客户分别为北京电视台、 中央电视台、云南电视台、北京盛世骄阳文化传播公司、深圳广播电影电视

集团， 去年同期五大客户分别为安徽电视台、上海聚力传媒技术有限公司、江苏广播电视 总台、河南电视台、上海东方娱乐传媒集团有限公司。 3、我们维持此前判断，认为公司过硬的电视剧精品质量为其强势发行奠定基础，各大卫视播放平台强强联合和组合模式不断创新为公司电视剧高收视创造条件。

In the first half of 2013, the TV drama business was steadily advancing, both in quality and quantity, and the overall quality of customers was improved.

In the first half of 2013, ***4 dramas were given issuance licenses, and 7 dramas have already started/completed shooting, pending issuance license.*** (1) The TV dramas broadcasted in the first half of the year include: "Innocent encounters reality" premiered in the three satellites TV of the East, Jiangsu, and Shenzhen, and won the national TV ratings at the same time; Xu Jizhou's "Thunder" was in Beijing, Anhui, Shenzhen, and Yunnan. Home TV broadcasted in the first round, and the ratings were slightly lower than expected. (2) During the reporting period, the four drama obtained the issuance license: “Wind and Thunder”, “Zero Cannon”, “Thunder”, “Romantic My Heart”. (3) There are 7 dramas that have not yet obtained the issuance license: "Golden Wars", "Year of Time" (formerly known as "Beijing Family"), "Behind Men", "Let's get married" (will be broadcast in CCTV -1 and Hunan Satellite TV during prime time in 2013), "Beautiful Contract", "Boiled Woman is Crazy", "Contract of Love". Among them, the return on investment on three of the dramas is over 90%;

The top five dramas contributed significantly to core income, and the overall quality of customers improved. (1) In the first half of the year, the top five film and television dramas of the company's main business income were "War of Thunder", "Zero Cannon House", "Never Desolate Number", "Different Brothers", "Good Times of Daughter-in-law", it accounted for 94.33% of the company's main business income during the same period. (2) ***From the customer's point of view, the overall quality of the company's customers has improved. In the first half of the year, the company's top five customers were Beijing TV, CCTV, Yunnan TV, Beijing Shengshi Sunshine Culture Communication Company and Shenzhen Radio, Film and Television Group. The top five customers in the same period last year were Anhui TV Station, Shanghai Juli Media Technology Co., Ltd. and Jiangsu Radio and TV Group. Taiwan, Henan TV Station, Shanghai Oriental Entertainment Media Group Co., Ltd.*** 3. We maintain our previous judgments and believe that the quality of the company's excellent TV drama quality lays the foundation for its strong issuance. **The company’s broadcast platform (TV station customers) will continue to improve,** and the combination mode is constantly innovating to create conditions for the company's growth.

Idiosyncratic Example 4: Beverages 2014

激励机制改革，将成为公司盈利能力提升的最大动力。公司实际控制人为河 北省衡水市国资委，由于激励机制长期不到位，严重影响了公司盈利能力。 根据公司 2013 年年报，公司全体董事、监事和高级管理人员的报酬合计只 有 168 万元。2013 年，公司毛利率仅为 51%，公司净利润率仅为 3.6%， 毛利率和净利率均在行业排名靠后。此次非公开增发，公司董事、监事和高 级管理人员、员工、优秀经销商均参与认购，同时引入两家战略投资者。随 着公司激励机制改革的完成，将释放管理层和员工积极性，公司盈利能力有 望得到大幅提升，业绩有较大弹性。

公司具有较强营销实力，渠道进一步下沉带来新的增长空间。去年 11 月， 公司管理层换届完成，营销公司总经理王占刚出任公司总经理。王占刚是业 内知名的营销人才，年仅 42 岁，其到任有望给公司带来新的活力。公司在 河北省具有绝对领先地位，省内竞争对手主要为山庄老酒和板城烧锅酒。未 来公司将河北市场渠道下沉到县级及以下市场，并在环河北市场实现地级市 的全面覆盖。今年 6 月底，公司与酒仙网签约，携河北近万家终端店入驻酒 仙网旗下“酒快到”O2O 平台。随着公司渠道进一步下沉、区域拓展和新的 营销平台建设，公司市场仍有较大提升空间。

The reform of incentive mechanism will become the biggest driving force for the company's profitability improvement. ***The actual controlling entity of the company is the State-owned Assets Supervision and***

Administration Commission of Hengshui City, Hebei Province. Because the incentive mechanism has not been in place for a long time, it has seriously affected the company's profitability. According to the company's 2013 annual report, the total remuneration of all directors, supervisors and senior management of the company is only 1.68 million yuan. In 2013, the company's gross profit margin was only 51%, the company's net profit margin was only 3.6%, and the gross profit margin and net profit margin were both lower than the industry. The non-public issuance, the company's directors, supervisors and senior management personnel, employees, outstanding distributors are involved in the subscription, while introducing two strategic investors. With the completion of the reform of the company's incentive mechanism, the management and employees' enthusiasm will be released. The company's profitability is expected to be greatly improved and its performance will be more flexible.

The company has strong marketing strength, and the channel further sinks to bring new growth space. In November last year, the management of the company was re-elected, and Wang Zhangang, general manager of the marketing company, became the general manager of the company. Wang Zhangang is a well-known marketing talent in the industry. He is only 42 years old, and his appointment is expected to bring new vitality to the company. ***The company has an absolute leading position in Hebei Province.*** The main competitors in the province are Shangzhuang Laojiu and Bancheng Cauldron. ***In the future, the company will lower the Hebei market channel to the county-level and below markets, and achieve full coverage of prefecture-level cities in the Hebei market.*** At the end of June this year, the company signed a contract with Jiuxian.com and brought nearly 10,000 terminal stores in Hebei to the O2O platform of Jiuxian.com. With the further sinking of the company's channels, regional expansion and the construction of new marketing platforms, there is still much room for improvement in the company's market.

Idiosyncratic Example 5: Medical 2015

金赛药业凭借其生长激素独家的水针剂 型，在我国儿童矮小症治疗市场占有 50%以上的份额，而长效生长激素 的用药频次从每天 1 次降为每周 1 次，更加方便患者，一个疗程用药比 水针剂型贵 1 倍，有望继续提升公司生长激素市场份额及利润。另一个 重磅品种人促卵泡激素已获得药品批准文号，有望于 2015 年下半年上市，

公司是国家基因工程药物的龙头企业，在研储备丰富：水痘减毒活疫苗在北京、山西等地开展了两针免疫活动，待全部数据采集完毕即可进行数据统计和总结临床工作；带状泡疹减毒活疫苗正在进行伦理审批， 临床研究即将开展；胸腺素 α1 临床完成，正在申报生产等待药监局受理；

百益制药已进行了艾塞那肽注射液生产现场动态核查、抽样及材料的提 交，并完成抽样样品的检测，注册进展顺利。

公司中药产品亦在逐步推进：PGQ 及其注射液项目已申请 II、III 期 临床；洋参果冠心片的 III 期临床工作已入组 300 余例；莽吉柿胶囊的 II 期临床工作正在进行受试者筛选；银花泌炎灵片增加适应症的 II 期临床 受试者已全部出组，正在数据录入及分析。

公司在研产品的进展较为顺利，这些基因工程药物及中药产品能够 为公司未来的发展提供新的增长点

With its exclusive water injection dosage form of growth hormone, Jinsai Pharmaceutical has a 50% share in the treatment market for children with short stature in China, and the frequency of long-acting growth hormone has been reduced from once a day to once a week, making it more convenient for patients. One course of treatment is 1 times more expensive than water injection, and it is expected to continue to increase the company's growth hormone market share and profits. Another major breed of human follicle stimulating hormone has obtained the drug approval number and is expected to be available in the second half of 2015.

The company is a leading enterprise in national genetic engineering drugs, and is rich in research reserves: live attenuated varicella vaccine has carried out two-needle immunization activities in Beijing and Shanxi, and data statistics and clinical work can be carried out after all data collection is completed; ***The live***

attenuated vaccine against rash is undergoing ethical approval, and clinical research is about to be carried out; thymosin α1 is clinically completed and is being declared for production and awaiting acceptance by the Food and Drug Administration;

Baiyi Pharmaceutical has carried out on-site dynamic verification, sampling and material submission of exenatide injection production, and completed the sampling of samples, and the registration progressed smoothly.

The company's traditional Chinese medicine products are also gradually advancing: PGQ and its injection project have applied for Phase II and III clinical trials; Phase III clinical work of ginseng fruit and coronary heart tablets has been enrolled in more than 300 cases; Phase II clinical work of Qiji Persimmon Capsule is underway Subject screening was performed; Phase II clinical subjects with Yinhua Biyanling tablets increased indications were all out of the group and were being entered and analyzed.

The company's research products are progressing smoothly. These genetic engineering drugs and traditional Chinese medicine products can provide new growth points for the company's future development.

Appendix III: Representative paragraphs from analyst reports with industry-specific topics

Industry-Specific Example 1: Fashion 2011

正装子行业增速放缓。公司原有的主导产品西装、衬衫属于传统梭织服装中 的正装。目前从中国梭织服装与针织服装产量增速看，2000-2010 年针织服装与梭织服装产量复合增速分别为 17.5%、13.6%，无论是针织服装产量还是增速都已远远超越梭织服装。从全国重点大型百货商场服装销售情况看，国内男士西装、衬衫、西裤销量均呈现下降趋势，2004-2009 年男士西装、衬衫、西裤销量复合增速分别为 7%、4%、5%。综合这些数据看，服饰休闲化是未来服装行业发展的一个大趋势，正装子行业将回到 10%以下的增速水平。

转型多品牌经营正当其时。公司在经营管理过程中，深知行业这一特性，因 此，着力打造报喜鸟中高端品牌形象，通过提升品牌美誉度、产品附加值带来的售价和毛利率的提升以及稳健的渠道扩张来实现增长。公司定位 2000 元以上的中高档西装，未来提价幅度带来毛利率提升的边际效益逐渐减弱，这些都决定了报喜鸟品牌大幅开店的可能性比较小。我们认为，公司已经充分考虑到这些因素，当下实行多品牌策略也是大势所趋，明智之举。

The growth rate of the garment industry is slowing down. The company's original leading product suits and shirts are formal attire in traditional woven garments. ***At present, from the growth rate of China's woven garments and knitwear production, the combined growth rate of knitted garments and woven garments in 2000-2010 is 17.5% and 13.6% respectively.*** Both the output of knitted garments and the growth rate have far exceeded the shuttle. Weaving clothing. ***From the sales situation of clothing in major national department stores, the sales volume of domestic men's suits, shirts and trousers showed a downward trend.*** In 2004-2009, the sales growth rate of men's suits, shirts and trousers was 7%, 4% and 5% respectively. . Based on these data, apparel leisure is a major trend in the future development of the apparel industry, and the apparel industry will return to the growth rate below 10%.

Transforming multi-brand operations is the right time. ***In the process of business management, the company knows the characteristics of the industry. Therefore, it strives to build a high-end brand image of the news, and enhances the brand reputation***, the increase in the price and gross profit margin brought by the added value of the product, and the steady expansion of the channel. Achieve growth. The company's positioning of mid-to-high-end suits of more than 2,000 yuan, the marginal benefit of the increase in gross profit margin will gradually weaken in the future price increase, which determines the possibility of the big-opening of the newspaper. We believe that the company has fully considered these factors, and ***the implementation of multi-brand strategy is also a general trend***, wise.

图 5：国内服装年产量（万件）

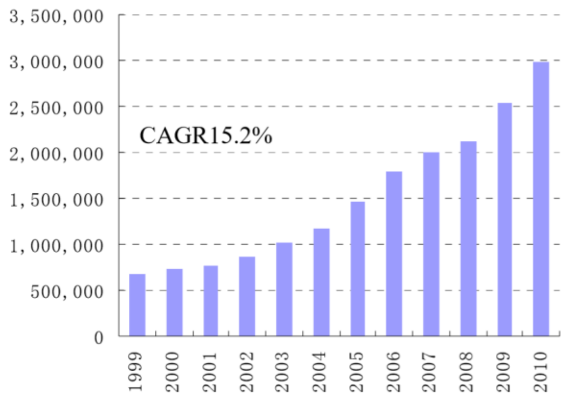


图 6：国内梭织、针织服装产量（万件）

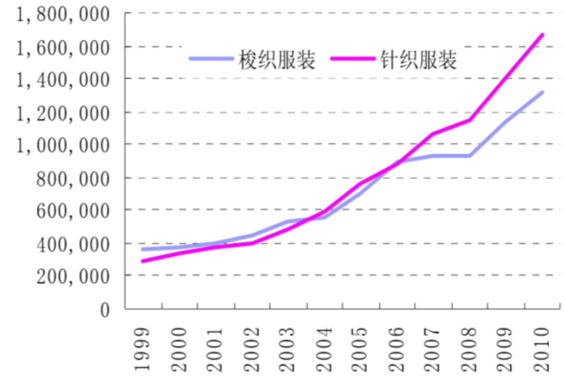


图 7：国内西服套装、衬衫产量（万件）

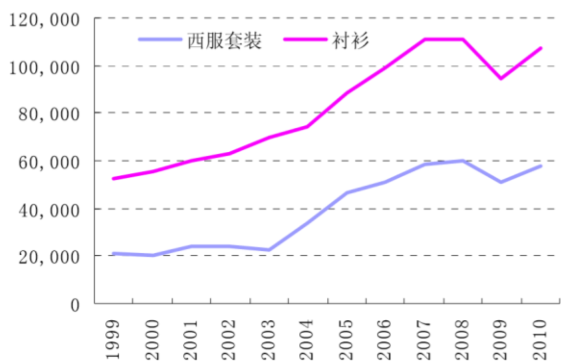
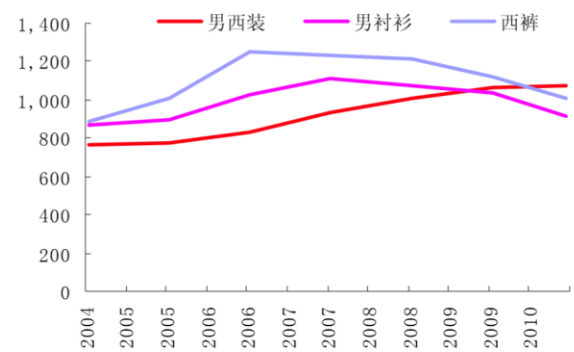


图 8：全国重点大型百货商场西服套装、衬衫销量（万件



Industry-Specific Example 2 Paper 2010

我们对公司的看法：白卡纸是公司的主要产品，公司未来目标是成为白卡纸的龙头企业，目前产能达到 75 万吨， 未来还将新建 75 万吨的产能。白卡纸的行业空间未来还比较大，这也是许多公司进入该行业的原因。就公司而言 我们看好其未来的前景。

我们认为纸业存在机会理由是：1、未来的销售旺季会带动行业销售的好转，2、行业中下游目前库存很少；3、九 月的落后产能淘汰我们认为可能超预期；4、除了铜版纸外，其他纸种在明年的 3 季度前并无大产能冲击。从供需 两个角度来看，未来纸业都趋好，我们维持前期的行业判断，纸业的价格、盈利 8 月是底，9 月开始回升。

Our view on the company: White cardboard is the company's main product. The company's future goal is to become the leading company of white cardboard, with a current production capacity of 750,000 tons, and a new capacity of 750,000 tons in the future. ***The industry space for white cardboard is still relatively large, which is why many companies have entered the industry. As far as the company is concerned, we are optimistic about its future prospects.***

We believe that the reasons for the opportunity in the paper industry are: 1. ***The future sales season will drive the improvement of the industry sales.*** *2.* ***The current middle and lower reaches of the industry are rarely in stock;*** 3. The backward production capacity in September is considered to be likely to exceed expectations; 4. ***In addition to the copper plate. Outside the paper, other paper types will not have a large capacity impact before the third quarter of next year. From the perspective of supply and demand, the paper industry is getting better in the future. We maintain the industry judgment in the early stage.*** The price and profit of the paper industry are at the end of August and start to rise in September.

Industry-Specific Example 3 Beverages 2014

主力产品受三公消费限制影响较小，公司积极布局大众酒。随着三公限酒的 影响，以及新型城镇化发展，我国白酒消费将由政商消费向大众消费转变。 公司高端酒十八酒坊受影响较大，但公司主力产品价位在 200 元以下，受影 响相对较小。2012 年底，公司成立了大众酒事业部，并于去年 6 月面向河 北省四大区域进行精准化招商，实现大众酒全省布局，推出的青花系列市场 反响较好。公司大众酒推出了八大系列，二十多个单品，价格覆盖 20-150 元。随着大众酒布局的完成，大众酒有望成为公司未来增长亮点。

The main products are less affected by the three public consumption restrictions, and the company actively distributes the public liquor. With the influence of the three public liquors and the development of new urbanization, China's liquor consumption will shift from political and commercial consumption to mass consumption. The company's high-end wine 18 winery was greatly affected, but the company's main product price is below 200 yuan, which is relatively small. At the end of 2012, the company established the Volkswagen Wine Division, and in June last year, it conducted precise investment promotion for the four major regions of Hebei Province, realizing the layout of the public liquor, and the blue-and-white series launched the market. The company's public liquor launched eight series, more than 20 single items, the price covers 20-150 yuan. ***With the completion of the public liquor layout, the public wine is expected to become the company's future growth highlights.***

Industry-Specific Example 4 Beverages 2010

莫高股份（600543）2010 年 2 月 3 日

我们看好国内葡萄酒行业的投资价值。公司具备“3-5 年内实现 10 亿元销售规模、冲击行业三甲”的实力：公司经营稳健，几年来葡 萄酒业务发展迅猛；公司的原料优势提供的不仅是原料，还是葡萄 酒中的精品。未来我们重点关注公司对外埠市场的拓展情况，这也 是公司目前最显著的“短”

We are optimistic about the investment value of the domestic wine industry. The company has the strength to achieve “1 billion yuan sales scale and impact the industry's top three in 3-5 years”: the company's operation is stable, and the wine business has developed rapidly in the past few years; the company's raw material advantage provides not only raw materials, but also the finest wines. In the future, we will focus on the expansion of the company's external market, which is also the company's most prominent "short board".

Industry-Specific Example 5 Fashion 2014

1、客户定位：定位于理性消费者，需要培养忠诚度

由于消费者较为理性，对产品设计和品质、产品形象理念和自身身份是否相符具有较 高的要求，该类品牌运营型企业必要在以下两方面投入巨大精力：一是注重产品品质和设 计，缔造经典形象；二是注重渠道终端形象、各种品牌推广活动及策略强化品牌形象。

除皮类产品外，女装、男装与同类竞争者相比，规模仍然偏小；从公司渠道分布看， 尤其是经销商渠道，主要分布在二三线城市；商场渠道受商场打折影响较大等问题，反映 出公司虽然数十年定位高端，但品牌影响力与白领、宝姿等国内品牌以及众多国际品牌相 比，仍存在一定差距。 未来随着高铁化时代到来，二三线市场品牌将会受到一定冲击；而随着国际品牌进一 步开拓中国市场，凯撒品牌影响力能否在上市后逐步有所提升，是决定公司中长期业绩增 速的根源。

Because consumers are more rational, and have high requirements for product design and quality, product image concept and their own identity, such brand-operated companies must invest a lot of energy in the following two aspects: First, focus on product quality and design, create The classic image; the

second is to focus on the channel terminal image, various brand promotion activities and strategies to strengthen the brand image.

In addition to leather products, women's wear and men's wear are still relatively small compared with competitors. From the perspective of company channel distribution, especially dealer channels, mainly distributed in second- and third-tier cities; shopping malls are affected by discounts on shopping malls. It reflects that although the company has positioned high-end for decades, its brand influence still has a certain gap compared with domestic brands such as white-collar and Baozi and many international brands. In the future, with the advent of the era of high-speed iron, the second and third-tier market brands will be affected by certain impacts; and as international brands further explore the Chinese market, whether the influence of Caesar brand will gradually increase after listing will determine the company's medium and long-term performance growth.

Appendix IV: Variable Definition	
WIdiosyncratic$_{i,j,t}$	Weight of idiosyncratic topics in percent for the report of analyst *i* of firm *j* in year *t*.
WOthers$_{i,j,t}$	Weight of other 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*.
$\text{Specialization}_{i,t}$	The number of different industries that the analyst *i* covers during year *t*.
$\text{Star}_{i,t}$	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 $\text{Analysts}_{j,t}$	The percentage of foreign analysts following firm *j* in year *t*
$\text{Loss}_{j,t}$	Indicator variable takes on the value of 1 if the firm *j* reports a loss in year *t*.
$\text{Foreign\_Ownership}_{j,t}$	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*
$\text{Foreign\_Ownership}_{j,t}$	Percentage of shares outstanding that is owned by foreign investors for firm *j* at the beginning of the calendar year *t*
Relative Forecast $\text{Error}_{i,j,t}$	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: $$SYNCH_{j,t} = \ln\left(\frac{R^2}{1-R^2}\right)$$ 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 idiosyncratic 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*.
$IdiosyncraticEarnings_{j,t}$	Similar to Ball and Brown (1967), we estimate the idiosyncratic component of earnings innovation using a three-step process: 1. For each firm-year observation, we estimate the following regression using 40 quarterly observations $\Delta ROA_{j,t-\delta} = \alpha_{i,t} + \beta_{j,t}\Delta ROA_{mk,t-\delta} + \gamma_{j,t}\Delta ROA_{ind,t-\delta} + \varepsilon$ 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. $E(\Delta ROA_{j,t}) = \widehat{\alpha_{j,t}} + \widehat{\beta_{j,t}}\Delta ROA_{mk,t} + \widehat{\gamma_{j,t}}\Delta ROA_{ind,t} + \varepsilon$ 3. The idiosyncratic earnings innovation is the absolute different between the actual $\Delta ROA_{j,t}$ and $E(\Delta ROA_{j,t})$, or: $IdiosyncraticEarnings_{j,t} = abs\left\vert\Delta ROA_{j,t} - E(\Delta ROA_{j,t})\right\vert$; i.e. the higher this number, the more idiosyncratic is the earnings innovation.
$Departure_{i,j,t}$	Indicator variable which takes on the value of 1 if anyone of firm *j*'s analysts has departed in the previous year.
$IdiosyncraticAnalyst_i$	Indicator variable takes on the value of 1 if the average $WIdiosyncratic_{i,j,t}$ across all of analyst *i*'s reports is ranked in the top quartile among all departing analysts.
$\# Analyst_{j,t}$	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: Idiosyncratic and Industry-specific Topics

Panel A - Number of idiosyncratic and industry-specific topics:

	Number of topics
Total	370
Industry-specific topics	60
Remaining topics	310
Idiosyncratic topics	103
Other topics	207

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

	Mean	**Median**	**Std.**	**Q1**	**Q3**
Industry specific	123.20	121.00	25.20	108.00	141.00
Idiosyncratic topics	48.48	40.00	16.00	20.00	71.00
Other topics	233.28	175.00	181.07	145.00	239.00

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

	Mean	**Median**	**Std.**	**Q1**	**Q3**
$WIndustry_{i,j,t}$	10.83%	4.25%	16.86%	1.64%	10.85%
$WIdiosyncratic_{i,j,t}$	44.93%	37.07%	32.64%	17.93%	69.34%
$WOthers_{i,j,t}$	44.22%	45.34%	32.41%	19.43%	71.68%

Panel D - Average number of topics in each report:

	Mean	**Median**	**Std.**	**Q1**	**Q3**
Industry specific	0.81	1.00	0.00	1.00	2.00
Idiosyncratic topics	3.22	3.00	2.00	2.00	4.00
Other topics	6.07	6.00	3.22	4.00	8.00

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 to which they are relevant. We classify the bottom 1/3 as idiosyncratic topics and the remaining topic as other 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: Idiosyncratic Information and Analyst Forecast Error

	(1)	(2)	(3)	(4)	(5)
			Relative Forecast Error$_{i,j,t}$		
WIdiosyncratic$_{i,j,t}$	-0.094***		-0.095***	-0.079***	-0.058***
	(-7.69)		(-7.57)	(-4.44)	(-3.37)
WIndustry$_{i,j,t}$		0.038*	0.036	0.022	0.023
		(1.71)	(1.41)	(0.91)	(0.94)
BM$_{j,t}$	0.049***	0.049***	0.049***	0.048***	0.014*
	(7.37)	(7.47)	(7.37)	(5.70)	(1.69)
Size$_{j,t}$	-0.026***	-0.025***	-0.026***	-0.022***	-0.016***
	(-12.05)	(-11.90)	(-12.07)	(-9.13)	(-6.20)
Volume$_{j,t}$	-0.008***	-0.007***	-0.008***	-0.014***	-0.020***
	(-3.27)	(-3.28)	(-3.27)	(-5.28)	(-7.07)
Stdret$_{j,t}$	0.003***	0.003**	0.003***	0.004***	0.008***
	(2.68)	(2.38)	(2.68)	(2.86)	(5.31)
Horizon$_{r,i,j,t}$	0.122***	0.122***	0.122***	0.123***	0.103***
	(26.08)	(26.03)	(26.06)	(26.11)	(17.84)
Experience_Firm$_{i,j,t}$	-0.008***	-0.008***	-0.008***	-0.007***	-0.008***
	(-4.49)	(-4.53)	(-4.49)	(-3.23)	(-3.76)
Experience$_{r,i,t}$	-0.009***	-0.009***	-0.009***	-0.008***	-0.084***
	(-6.72)	(-6.57)	(-6.72)	(-3.74)	(-6.83)
Brokersize_Analyst$_{i,t}$	-0.015	-0.009	-0.015	0.024	-0.264*
	(-0.31)	(-0.18)	(-0.31)	(0.19)	(-1.84)
Brokersize_Firms$_{i,t}$	-0.026***	-0.027***	-0.025***	-0.063***	0.037**
	(-3.57)	(-3.73)	(-3.56)	(-5.51)	(2.08)
Institutions_share$_{j,t}$	0.002***	0.002***	0.002***	0.002***	0.002***
	(5.71)	(5.82)	(5.71)	(4.63)	(4.66)
Specialization$_{i,t}$	-0.003***	-0.003***	-0.003***	-0.004**	-0.003
	(-3.62)	(-3.64)	(-3.62)	(-2.47)	(-1.62)
Star$_{i,t}$	-0.001	-0.000	-0.001	0.006	0.005
	(-0.09)	(-0.05)	(-0.09)	(0.66)	(0.54)
% Foreign Analysts$_{j,t}$	0.062	0.051	0.062	0.044	0.002
	(1.19)	(0.97)	(1.19)	(0.69)	(0.03)
Loss$_{j,t}$	-0.025***	-0.026***	-0.025***	-0.017**	-0.017**
	(-3.86)	(-3.95)	(-3.86)	(-2.12)	(-2.07)
Distance$_{i,j,t}$	0.009	0.006	0.009	0.011	0.010
	(1.05)	(0.67)	(1.05)	(1.27)	(1.15)
Foreign_Ownership$_{j,t}$	0.003*	0.003*	0.003*	0.003*	0.003
	(1.71)	(1.76)	(1.71)	(1.72)	(1.55)
Pages	0.000	0.000	0.000	0.000	0.000
	(1.21)	(1.07)	(1.21)	(0.24)	(0.41)
Fixed Effects	Firm	Firm	Firm	Firm, Analyst	Firm, Analyst, Year
Std. Cluster	Firm	Firm	Firm	Firm	Firm
Observations	87332	87332	87332	87332	87332

Adjusted R^2	0.075	0.074	0.075	0.087	0.088

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. $\frac{AFE_{i,j,t} - AvgAFE_{j,t}}{AvgAFE_{j,t}}$. $WIdiosyncratic_{i,j,t}$ represents the sum of all idiosyncratic topic weight, $WIndustry_{i,j,t}$ represents the sum of all industry-specific topic weight.

Table 4: Validation for our proxy of idiosyncratic information

	(1)	(2)	(3)	(4)
	$Synch_{j,t}$		$IdiosyncraticEarnings_{j,t}$	
$Avg\_Idiosyncratic_{j,t}$	-0.145^{**}		0.010^{**}	
	(-3.10)		(3.41)	
$Avg\_Industry_{j,t}$		0.294^{*}		-0.001
		(2.26)		(-0.23)
$Fund\_Corr_{j,t}$	0.018	0.019	-0.013^{***}	-0.013^{***}
	(0.34)	(0.38)	(-4.22)	(-4.13)
$\ln(Herf_{j,t})$	0.014	0.014	0.005^{**}	0.005^{**}
	(0.87)	(0.82)	(2.89)	(2.86)
$STD\_ROA_{j,t}$	-0.080^{**}	-0.082^{**}	0.002	0.002
	(-3.02)	(-3.04)	(1.72)	(1.86)
$\ln(NIND_{j,t})$	0.019	0.018	0.005	0.005
	(0.24)	(0.23)	(1.44)	(1.42)
$Size_{j,t}$	-0.380^{***}	-0.386^{***}	0.017^{***}	0.018^{***}
	(-6.82)	(-6.99)	(6.11)	(6.15)
Fixed Effects	Firm, Year	Firm, Year	Firm, Year	Firm, Year
Std. Cluster	Firm, Year	Firm, Year	Firm, Year	Firm, Year
Observations	10384	10384	10384	10384
Adjusted R^2	0.336	0.336	0.573	0.572

t statistics in parentheses

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

The above table include 10,384 firm-year observations. $Synch_{j,t}$ is a firm-specific measures of return synchronicity using the methodology outlined in Piotroski and Roulstone (2004). $IdiosyncraticEarnings_{j,t}$ is the idiosyncratic component of earnings innovation estimated using the procedure outlined in Ball and Brown (1967). *$Avg\_Idiosyncratic_{j,t}$* is the average idiosyncratic 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

Panel A: Idiosyncratic information

	(1)	(2)	(3)
	Relative Forecast Error$_{i,j,t}$		
$WIdiosyncratic_{i,j,t}$×$HighPolitical1_{j,t}$	-0.039***		
	(-2.95)		
$HighPolitical1_{i,j,t}$	-0.003		
	(-0.94)		
$WIdiosyncratic_{i,j,t}$× $HighPolitical2_{j,t}$		-0.031**	
		(-2.56)	
$HighPolitical2_{i,j,t}$		-0.004	
		(-1.17)	
$WIdiosyncratic_{i,j,t}$× $HighPolitical3_{j,t}$			-0.046**
			(-2.42)
$HighPolitical3_{i,j,t}$			-0.003
			(-1.16)
$WIdiosyncratic_{i,j,t}$	-0.071***	-0.079***	-0.083***
	(-4.63)	(-5.92)	(-4.41)
BM$_{j,t}$	0.048***	0.048***	0.051***
	(7.29)	(7.30)	(6.72)
Size$_{j,t}$	-0.026***	-0.026***	-0.012***
	(-12.08)	(-12.01)	(-4.85)
Volume$_{j,t}$	-0.008***	-0.008***	-0.009***
	(-3.37)	(-3.67)	(-3.40)
Stdret$_{j,t}$	0.003***	0.003**	0.008***
	(2.61)	(2.20)	(5.22)
Horizon$_{r,i,j,t}$	0.122***	0.123***	0.103***
	(26.19)	(26.21)	(27.46)
Experience_Firm$_{i,j,t}$	-0.008***	-0.008***	-0.009***
	(-4.40)	(-4.37)	(-5.74)
Experience$_{r,i,t}$	-0.009***	-0.010***	-0.004***
	(-6.59)	(-6.96)	(-3.21)
Brokersize_Analyst$_{i,t}$	-0.009	-0.017	-0.043
	(-0.19)	(-0.36)	(-0.96)
Brokersize_Firms$_{i,t}$	-0.027***	-0.026***	0.002
	(-3.72)	(-3.66)	(0.33)
Institutions_share$_{j,t}$	0.002***	0.002***	0.002***
	(5.68)	(5.65)	(4.34)
Specialization$_{i,t}$	-0.003***	-0.003***	-0.003***
	(-3.70)	(-3.65)	(-3.64)
Star$_{i,t}$	-0.001	-0.001	-0.002
	(-0.09)	(-0.13)	(-0.27)
% Foreign Analysts$_{j,t}$	0.065	0.067	-0.176***
	(1.24)	(1.28)	(-2.89)
Loss$_{j,t}$	-0.026***	-0.026***	-0.035***

	(-3.87)	(-4.00)	(-4.28)
$Distance_{i,j,t}$	0.009	0.008	0.009
	(1.03)	(1.00)	(1.21)
$Foreign\_Ownership_{j,t}$	0.003^{*}	0.003^{*}	0.002
	(1.67)	(1.73)	(0.93)
Fixed Effects	Firm	Firm	Firm
Std. Cluster	Firm	Firm	Firm
Observations	87332	87332	87332
Adjusted R^2	0.076	0.076	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. $\frac{AFE_{i,j,t}-AvgAFE_{j,t}}{AvgAFE_{j,t}}$., *WIdiosyncratic*$_{i,j,t}$ represents the sum of all idiosyncratic 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 5: Political influence at the firm-level

Panel B: Industry-specific information

	(1)	(2)	(3)
	Relative Forecast Error$_{i,j,t}$		
$WIndustry_{i,j,t} \times HighPolitical1_{j,t}$	0.026		
	(1.38)		
$HighPolitical1_{i,j,t}$	0.003		
	(0.92)		
$WIndustry_{i,j,t} \times HighPolitical2_{j,t}$		0.011	
		(0.47)	
$HighPolitical2_{i,j,t}$		0.004	
		(1.37)	
$WIndustry_{i,j,t} \times HighPolitical3_{j,t}$			-0.015
			(-0.26)
$HighPolitical3_{i,j,t}$			0.051
			(1.38)
$WIndustry_{i,j,t}$	0.040	0.066***	0.079**
	(1.43)	(2.60)	(2.24)
BM$_{j,t}$	0.049***	0.048***	0.051***
	(7.41)	(7.36)	(6.78)
Size$_{j,t}$	-0.025***	-0.025***	-0.011***
	(-11.86)	(-11.85)	(-4.61)
Volume$_{j,t}$	-0.008***	-0.009***	-0.009***
	(-3.41)	(-3.77)	(-3.42)
Stdret$_{j,t}$	0.003**	0.002*	0.008***
	(2.29)	(1.88)	(4.93)
Horizon$_{r,i,j,t}$	0.122***	0.123***	0.102***
	(26.15)	(26.19)	(27.40)
Experience_Firm$_{i,j,t}$	-0.008***	-0.008***	-0.009***
	(-4.42)	(-4.36)	(-5.78)
Experience$_{r,i,t}$	-0.009***	-0.010***	-0.004***
	(-6.51)	(-6.85)	(-3.03)
Brokersize_Analyst$_{i,t}$	-0.004	-0.013	-0.036
	(-0.08)	(-0.26)	(-0.82)
Brokersize_Firms$_{i,t}$	-0.028***	-0.027***	0.000
	(-3.87)	(-3.83)	(0.07)
Institutions_share$_{j,t}$	0.002***	0.002***	0.002***
	(5.75)	(5.73)	(4.44)
Specialization$_{i,t}$	-0.003***	-0.003***	-0.003***
	(-3.70)	(-3.67)	(-3.68)
Star$_{i,t}$	-0.001	-0.001	-0.002
	(-0.09)	(-0.15)	(-0.26)
% Foreign Analysts$_{j,t}$	0.056	0.057	-0.184***
	(1.08)	(1.10)	(-3.05)
Loss$_{j,t}$	-0.026***	-0.027***	-0.035***

	(-3.99)	(-4.10)	(-4.34)
$Distance_{i,j,t}$	0.005	0.005	0.005
	(0.63)	(0.62)	(0.70)
$Foreign\_Ownership_{j,t}$	0.003^{*}	0.003^{*}	0.002
	(1.76)	(1.77)	(0.97)
Fixed Effects	Firm	Firm	Firm
Std. Cluster	Firm	Firm	Firm
Observations	87332	87332	87332
Adjusted R^2	0.075	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. $\frac{AFE_{i,j,t}-AvgAFE_{j,t}}{AvgAFE_{j,t}}$. *WIndustry*$_{i,j,t}$ represents the sum of all industry-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
Panel A: Idiosyncratic information

	(1)	(2)	(3)	(4)
	Relative Forecast Accuracy$_{i,j,t}$			
$WIdiosyncratic_{i,j,t} \times Turnover_{j,t}$	-0.038***			
	(-4.00)			
$Turnover_{j,t}$	0.009**			
	(2.00)			
$WIdiosyncratic_{i,j,t} \times Policy_{j,t}$		-0.101***		
		(-3.92)		
$Policy_{j,t}$		0.049***		
		(2.83)		
$WIdiosyncratic_{i,j,t} \times HighGovInt_{j,t}$			-0.057***	
			(-2.99)	
$HighGovInt_{j,t}$			-0.021	
			(-1.61)	
$WIdiosyncratic_{i,j,t} \times LowPptRight_{j,t}$				-0.055***
				(-2.75)
$LowPptRight_{j,t}$				-0.010
				(-1.57)
$WIdiosyncratic_{i,j,t}$	-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. $\frac{AFE_{i,j,t}-AvgAFE_{j,t}}{AvgAFE_{j,t}}$. *WIdiosyncratic*$_{i,j,t}$ represents the sum of all idiosyncratic topic weight. *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 province 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 6: Political influence at the state-level
Panel B: Industry-specific information

	(1)	(2)	(3)	(4)
	Relative Forecast $Accuracy_{i,j,t}$			
$WIndustry_{i,j,t} \times Turnover_{j,t}$	0.080***			
	(4.58)			
$Turnover_{j,t}$	-0.007			
	(-1.04)			
$WIndustry_{i,j,t} \times Policy_{j,t}$		0.029		
		(0.58)		
$Policy_{j,t}$		0.005		
		(0.31)		
$WIndustry_{i,j,t} \times HighGovInt_{j,t}$			0.104***	
			(2.83)	
$HighGovInt_{j,t}$			-0.058	
			(-1.40)	
$WIndustry_{i,j,t} \times LowPptRight_{j,t}$				0.076*
				(1.73)
$LowPptRight_{j,t}$				-0.082
				(-1.33)
$WIndustry_{i,j,t}$	-0.015	0.009	-0.022	-0.011
	(-0.58)	(0.21)	(-0.85)	(-0.35)
$BM_{j,t}$	-0.113***	-0.113***	-0.112***	-0.110***
	(-9.06)	(-9.08)	(-8.98)	(-8.79)
$Size_{j,t}$	0.046***	0.049***	0.050***	0.051***
	(6.98)	(7.38)	(7.29)	(7.31)
$Volume_{j,t}$	-0.026***	-0.026***	-0.027***	-0.026***
	(-12.04)	(-11.99)	(-11.99)	(-11.57)
$Stdret_{j,t}$	-0.008***	-0.008***	-0.009***	-0.010***
	(-3.63)	(-3.36)	(-3.83)	(-4.03)
$Horizon_{r,i,j,t}$	0.004***	0.003***	0.003**	0.004***
	(3.43)	(2.83)	(2.32)	(2.62)
$Experience\_Firm_{i,j,t}$	0.122***	0.122***	0.123***	0.121***
	(26.18)	(26.17)	(26.28)	(26.03)
$Experience_{r,i,t}$	-0.008***	-0.008***	-0.008***	-0.008***
	(-4.51)	(-4.46)	(-4.46)	(-4.55)
$Brokersize\_Analyst_{i,t}$	-0.010***	-0.010***	-0.008***	-0.013***
	(-6.91)	(-6.79)	(-5.63)	(-8.55)
$Brokersize\_Firms_{i,t}$	-0.020	-0.016	-0.028	-0.020
	(-0.42)	(-0.33)	(-0.59)	(-0.42)
$Institutions\_share_{j,t}$	-0.024***	-0.026***	-0.023***	-0.023***
	(-3.40)	(-3.57)	(-3.24)	(-3.20)
$Specialization_{i,t}$	0.002***	0.002***	0.002***	0.002***
	(5.66)	(5.68)	(5.55)	(5.87)
$Star_{i,t}$	-0.003***	-0.003***	-0.003***	-0.003***
	(-3.53)	(-3.62)	(-3.60)	(-3.61)

% Foreign Analysts$_{j,t}$	-0.001	-0.001	-0.001	-0.001
	(-0.17)	(-0.14)	(-0.16)	(-0.16)
Loss$_{j,t}$	0.073	0.069	0.093*	0.052
	(1.38)	(1.31)	(1.69)	(0.91)
Foreign_Ownership$_{j,t}$	-0.025***	-0.025***	-0.026***	-0.029***
	(-3.74)	(-3.82)	(-3.84)	(-3.98)
Fixed Effects	0.003*	0.003*	0.003	0.003*
Std. Cluster	(1.66)	(1.67)	(1.49)	(1.67)
Observations	87332	87332	87332	87332
Adjusted R^2	0.075	0.074	0.075	0.074

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. $\frac{AFE_{i,j,t}-AvgAFE_{j,t}}{AvgAFE_{j,t}}$. *WIndustry*$_{i,j,t}$ represents the sum of all industry-specific topic weight. *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 province 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: Analyst Departure and Stock Return Synchronicity

Panel A – Analyst turnover frequency

	n	%
Unique # analysts in our sample	3175	100%
Analyst that turnover during our sample period	1128	35.5%
Analyst that turnover in 2010	*305*	
Analyst that turnover in 2011	*252*	
Analyst that turnover in 2012	*267*	
Analyst that turnover in 2013	*304*	

Panel B – Idiosyncratic analyst vs. non-idiosyncratic analyst characteristics

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

	(1)	(2)
	$Synch_{j,t}$	$IdiosyncraticEarnings_{j,t}$
$Departure_{i,t} \times IdiosyncraticAnalyst_i$	0.109^{**}	0.002
	(3.00)	(0.92)
$IdiosyncraticAnalyst_i$	0.006	-0.002
	(0.22)	(-0.89)
$Departure_{i,t}$	-0.110	-0.003
	(-0.97)	(-1.16)
$Fund\_Corr_{j,t}$	0.085	-0.023^{*}
	(1.59)	(-1.98)
$In(Herf_{j,t})$	0.019^{*}	-0.000
	(1.95)	(-0.26)
$STD\_ROA_{j,t}$	-0.348	-1.390^{*}
	(-1.12)	(-2.05)
$In(NIND_{j,t})$	-0.000	0.000
	(-0.71)	(0.00)
$Size_{j,t}$	-0.331^{***}	0.014^{**}
	(-8.62)	(2.44)
Fixed Effects	Firm, Year	
Std. Cluster	Firm, Year	
Observations	11357	11357
Adjusted R^2	0.376	0.595

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

The table above includes 11,357 firm-year observations. The unit of analysis is firms with a departing analyst (*i*)- pre-/post-(*t*). Only departure of idiosyncratic and non-idiosyncratic analysts are included. $Synch_{j,t}$ is a idiosyncratic measures of return synchronicity using the methodology outlined in Piotroski and Roulstone (2004). $IdiosyncraticEarnings_{j,t}$ is the idiosyncratic 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. $IdiosyncraticAnalyst_i$ is an indicator variable which takes the value of 1 if the departing analyst's $WIdiosyncratic_{i,j,t}$ is ranked in the top quartile.